

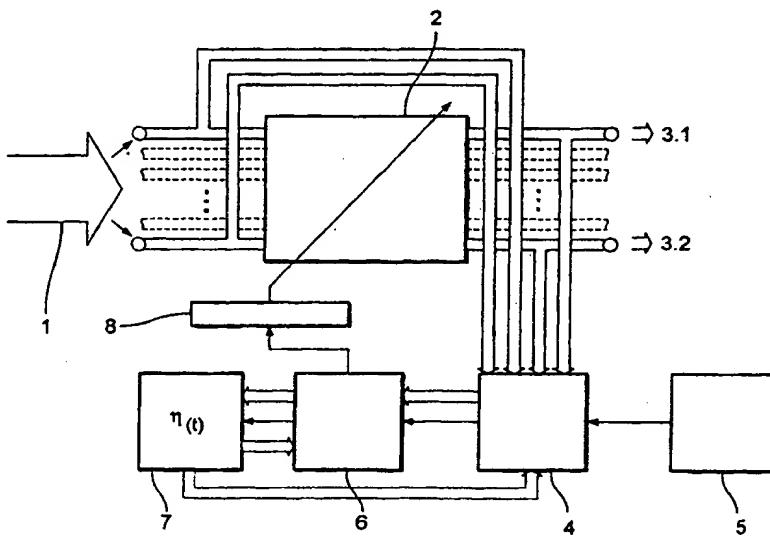
(51) Internationale Patentklassifikation 6 : G06F 15/80	A1	(11) Internationale Veröffentlichungsnummer: WO 98/25214 (43) Internationales Veröffentlichungsdatum: 11. Juni 1998 (11.06.98)
(21) Internationales Aktenzeichen: PCT/DE97/02902 (22) Internationales Anmeldedatum: 3. Dezember 1997 (03.12.97)	(81) Bestimmungsstaaten: JP, US, europäisches Patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(30) Prioritätsdaten: 196 52 336.2 3. Dezember 1996 (03.12.96) DE	Veröffentlicht <i>Mit internationalem Recherchenbericht. Vor Ablauf der für Änderungen der Ansprüche zugelassenen Frist. Veröffentlichung wird wiederholt falls Änderungen eintreffen.</i>	
(71) Anmelder (für alle Bestimmungsstaaten ausser US): GMD – FORSCHUNGSZENTRUM INFORMATIONSTECHNIK GMBH [DE/DE]; Schloss Birlinghoven, D-53757 Sankt Augustin (DE).		
(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): MÜLLER, Klaus-Robert [DE/DE]; Kochhannstrasse 24, D-10249 Berlin (DE). ZIEHE, Andreas [DE/DE]; Kirchplatz 14, D-14947 Ruhlsdorf (DE). MURATA, Noboru [JP/JP]; Hirosawa 2-1, Wako-shi, Saitama 351-01 (JP).		
(74) Anwalt: CHRISTIANSEN, Henning; Pacelliallee 43/45, D-14195 Berlin (DE).		

(54) Title: METHOD AND DEVICE FOR NON-STATIONARY SEPARATION OF SIGNALS OF VARIOUS SOURCES

(54) Bezeichnung: VERFAHREN UND VORRICHTUNG ZUR NICHTSTATIONÄREN TRENNUNG VON SIGNALEN UNTERSCHIEDLICHER QUELLEN

(57) Abstract

Method and device for non-stationary separation of signals of various sources or separation of at least one signal background, comprising a demixing phase with a switching matrix allocating, in a controlled manner, a number of signal inputs with mixing signal components to various outputs on evaluation elements by using an adaptive adding or multiplying algorithm. The adaptive algorithm can be variably controlled by using a variable learning rate, wherein the learning rate is continuously adapted by inclusion of an algorithm which comprises the following steps: i) detection of a differentiable function or at least two matrixes from the signal, the separated signal and the learning rate, which evaluates the separated signal; ii) adaptation of the switching matrix based on results of (i) and the learning rate, iii) adaptation of the learning rate based on results of (i) and (ii).



(57) Zusammenfassung

Verfahren und Vorrichtung zur nichtstationären Trennung von Signalen unterschiedlicher Quellen oder Trennung mindestens eines Signals von einem Signalhintergrund, mit einer Entmischungsstufe, die eine Schaltmatrix aufweist, die eine Anzahl von Mischungssignalanteilen aufweisenden Signaleingängen unterschiedlichen Ausgängen über Bewertungselemente jeweils additiv oder multiplikativ über einen adaptiven Algorithmus steuerbar zuordnet, wobei der adaptive Algorithmus unter Nutzung einer variablen Lernrate veränderbar angesteuert wird, wobei die Lernrate kontinuierlich unter Einschluß eines Algorithmus angepaßt wird, welcher die Schritte (i) Ermittlung einer differenzierbaren Funktion oder mindestens zweier Matrizen aus dem Signal, dem getrennten Signal und der Lernrate, die das getrennte Signal bewertet bzw. bewerten, (ii) Adaptation der Schaltmatrix auf der Basis des Ergebnisses von (i) sowie der Lernrate, (iii) Adaptation der Lernrate auf der Basis der Ergebnisse von (i) und (ii), umfaßt.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		

**Verfahren und Vorrichtung zur nichtstationären Trennung von
Signalen unterschiedlicher Quellen**

Beschreibung

Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur nichtstationären Quellentrennung (Quellenvereinzelung) gemäß dem Oberbegriff des Anspruchs 1.

Die Quellentrennung ist ein sowohl theoretisch interessantes als auch für Anwendungen hochrelevantes Gebiet. Die Extraktion eines Signals aus einem Signalgemisch oder aus einem Signalhintergrund aufgrund seiner unabhängigen Merkmale findet sich nicht nur in primären sensorischen Arealen des Gehirns, sondern spielt auch für die technische Signalverarbeitung eine zentrale Rolle als Grundprinzip der Vorverarbeitung (J.J. Hopfield in *Proc. natl. Acad. of Sciences USA*, Bd. 88, S. 6462 (1991); H.B. Barlow et al., in „The computing neuron“, hrsg. von R. Durbin et al., S. 54, Addison Wesley 1989; C. Jutten et al. in *Signal Processing*, Bd. 10, 24, S. 1 (1991); P. Comon in *Signal Processing*, Bd. 36, S. 287 (1994); T. Bell et al. in *Neural Comp.*, Bd 7, S. 1129 (1995a)).

So kommt es in vielen Problemen der Signalverarbeitung vor, daß das relevante Signal stark mit Rauschen, mit einem oder 20 sogar mit mehreren anderen Signalen vermischt ist. Verrauschte oder vermischte Signale lassen sich nur schwierig erkennen und weiterverarbeiten, daher sind Algorithmen, die in kontrollierter Weise Signale und Störung trennen, von großer Wichtigkeit. Eine lineare Mischung wäre beispielsweise

$$\bar{I} = M\bar{s} \quad (1)$$

wobei I das gemessene bekannte Signal ist und sowohl das Mischungsverhältnis M als auch die Bestandteile des Orig

nalsignals s unbekannt sind. Konvolutionen oder nichtlineare Mischungen können als Verallgemeinerungen von Gl.(1) definiert werden.

Klassische Methoden, die auf Filtern basieren, setzen vor-
5 aus, daß sich Spektren von Signal und Rauschen stark unterscheiden.

Im allgemeinen hat man aber weder a-priori Wissen über die Wahrscheinlichkeitsverteilung der Signale noch über die Störung. Dabei hat es sich als zweckmäßig erwiesen, die
10 Wahrscheinlichkeitsdichten der vermischten Prozesse blind (datengetrieben) zu schätzen. Für die Quellentrennung versuchte man dabei Architekturen zu finden, welche die Transformation zwischen Eingangs- und Ausgangssignalen maximieren oder aber untereinander minimal korrelierte Ausgabesig-
15 gnale zu erhalten (G. Decco et al., „An information-theoretic approach to neural computing“, Springer 1996). Üblicherweise wird dazu die Entropie der einzelnen Ausgaben approximiert.

Dies geschieht beispielsweise durch eine Edgeworth oder ei-
20 ne Gram-Charlier Entwicklung der Momente (cf. Jutten et al 1991; Common 1994; S. Amari et al., in „Advances in Neural Information Processing Systems 8 (NIPS 95)“, hrsg. von D.S. Touretzky et al., MIT Press 1995). Bell et al. (1995a) ver-
wenden eine informationstheoretisch motivierte Anti-Hebb-
25 Regel, um sich die notwendigen Momente höherer Ordnung zu erzeugen. Eine weitere sehr elegante Methode zur Entmi-
schung ist ein von Molgedy und Schuster vorgestellter li-
nearer Algorithmus. Hier werden nicht nur gleichzeitige,
sondern auch zeitversetzte Kreuzkorrelationen der Ausgaben
30 minimiert und so kann durch Zeitversetzung die Berücksich-

tigung von Korrelationen höherer Ordnung vermieden werden (L. Molgedey und H.G. Schuster, in *Phys. Rev. Lett.*, Bd. 72, S. 3634 (1994)).

- Quellentrennungsalgorithmen werden u.a. zur Analyse von EEG
5 Signalen eingesetzt (cf. S. Makeig et al., in „*Advances in Neural Information Processing Systems 8 (NIPS 95)*“, hrsg. von D.S. Touretzky et al., S. 145, MIT Press 1995; Bell et al. 1995a) und sind auch für die Spracherkennung von großem Nutzen (T. Bell et al. in „*Proc. of ICASSP '95*“, 1995b).
- 10 Man unterscheidet zwischen der Quellentrennung, in der (a) stationäre und (b) instationäre Mischungen vorliegen. Während Fall (a) Stand der Technik ist, ist der nichtstationäre Fall (b) bisher technisches Neuland. In der Praxis liegen normalerweise instationäre Mischungen vor - Fall (b).
15 Es gibt wenige Algorithmen, die potentiell für den nichtstationären Fall (b) von Nutzen sein könnten. Ein Algorithmus für den nichtstationären Fall muß nämlich ein On-line-Algorithmus sein, der schnell genug eine Änderung der Mischung erkennen und ihr folgen kann.
- 20 Es gibt auch eine Reihe von On-line-Algorithmen, die Klassifikations- oder Regressionsanwendungen ihren Einsatz finden (Saad et al., *Workshop at NIPS '95*, siehe <http://neural-server.aston.ac.uk/nips95/workshop.html>, und darin enthaltene Quellen (1995)). Einige dieser On-line-
25 Verfahren (C. Darken et al., in „*Advances in Neural Information Processing Systems 4 (NIPS 91)*“, hrsg. von J. E. Moody et al., San Mateo CA: Morgan Kaufmann, S. 1009 (1992)) sind sogar in der Lage, eine stationäre Lösung - Fall (a) - gut zu approximieren, indem sie zwei Phasen der Konvergenz
30 durchlaufen, in der die Lernrate η unterschiedlich angepaßt

- 5 -

wird. Es wird dabei ein Gradientenabstieg in den Parametern w

$$\hat{w}_{t+1} = \hat{w}_t - \eta_t C(\hat{w}_t) \frac{\partial}{\partial w} l(x_{t+1}m, y_{t+1}; \hat{w}_t) \quad (2)$$

benutzt, wobei $l(x_{t+1}, y_{t+1}; \hat{w}_t)$ die Fehler- bzw. Likelihood-Funktion ist. Der Index t beschreibt die Zeit bzw. die schon gesehene Anzahl der Daten, C ist eine Matrix, die unterschiedlich gewählt werden kann.

In einer ersten Phase wird die Lernrate η konstant gehalten:

$$10 \quad \eta = \text{konstant}, \quad (3)$$

um nachfolgend in einer zweiten Phase, sobald der stochastische Suchprozess (Gradientenabstieg aus Gl.(2)) in die Nähe des gesuchten Minimums kommt, reduziert zu werden:

$$15 \quad \eta = \frac{c}{t} \quad (4)$$

Im Jargon spricht man hier von einem "Abkühlen" der Lernrate.

Je mehr Daten das Lernverfahren sieht (je größer also der Index t aus Gl.(4) wird), desto geringer wird die Lernrate η und desto geringer die Suchgeschwindigkeit im Phasenraum.

Ein herkömmlicher On-Line-Algorithmus wird also nach gewisser Zeit in Phase 1 in die zweite Phase übergehen und zur Lösung konvergieren.

Ändert sich die Statistik der Daten, zum Beispiel im Falle der Quellentrennung das Mischungsverhältnis der Signale,

dann wird das herkömmliche On-Line-Verfahren der veränderten Lösung nur mit winzigen (und immer kleiner werdenden) Suchschritten (cf. Gl.4) hinterherreilen, da es nicht mehr zur ersten Phase - der konstanten großen Lernrate - zurückkehren kann. Das bedeutet, daß die neue Lösung mit veränderter Datenstatistik gar nicht erreicht wird, somit das On-line-Verfahren trotz Adaptivität in der Praxis völlig nutzlos ist.

Die oben beschriebenen On-line-Algorithmen sind in der Praxis nicht schnell genug in der Lage, eine Nichtstationarität zu erkennen und zu verfolgen, da sie nicht automatisch aus der zweiten Phase in die erste zurückkehren können, somit sind die beschriebenen On-line-Algorithmen für den instationären Fall (b) trotz Adaptivität nutzlos.

Betrachtet man On-line-Lernen mit konstanter Lernrate, so kann man dabei einer Änderung der Statistik folgen, jedoch ist es relativ langsam und aus systematischen Gründen nicht in der Lage, die Lösung sauber zu approximieren. Denn die Varianz oder Abweichung von der Optimallösung ist immer von der Größenordnung der Lernrate abhängig. So besteht bei konstanter Lernrate das Problem, entweder mit großer Lernrate einer Änderung der Statistik schnell zu folgen, aber gleichzeitig auch immer weit (Größenordnung der Lernrate) von der Lösung weg zu sein, oder mit kleiner Lernrate eine Statistikänderung zu langsam zu verfolgen, aber gleichzeitig dafür der Lösung etwas näher zu sein.

Da der Batch-Algorithmus alle Daten zur Schätzung nimmt, ist er im instationären Fall immer weit weg von der optimalen Lösung, da er den Mittelwert der sich zeitlich verän-

dernden Lösung berechnet und sich somit als relativ nutzlos erweist.

Nur eines dieser Verfahren ist in der Lage, Nichtstationarität zu erkennen und sich schnell anzupassen. Sompolinski et al 1995 definierten diesen sogenannten "learning of learning" Algorithmus. Der "learning of learning" Algorithmus ist in der Lage, zu detektieren, wann sich die Statistik der Daten verändert und automatisch in die erste Phase der stochastischen Suche mit konstanter (großer) Lernrate η zurückzuschalten. Nach schnellem Konvergieren bis in Lösungs-nähe (erste Phase), wird die Lernrate wiederum "abgekühlt" (zweite Phase cf. Gl.(4)). Leider ist das "learning of learning" Verfahren nur für ein einfaches Perceptron mit einer diskreten Fehlerfunktion definiert und ist daher in der von H. Sompolinski et al. in „Neural Networks: The Statistical Mechanics Perspective“, hrsg. von J.-H. Oh et al., S.105 Singapore: World Scientific (1995) angegebenen Form nicht allgemein anwendbar. Insbesondere sind Quellentrennungsprobleme kontinuierlich und nicht diskret und es kann keine explizite Fehlerfunktion definiert werden.

Zusammenfassend bestehen also folgende Nachteile der bekannten Verfahren:

1. Herkömmliche adaptive On-line-Algorithmen können nicht effizient und schnell mit Nichtstationarität der Datenstatistik umgehen. Man kann zeigen, daß sie die veränderte Optimallösung nicht erreichen können.
2. On-line-Algorithmen mit fester Lernrate sind systematisch entweder ungenau oder langsam.

3. Batch Verfahren approximieren den Mittelwert aller zwischenzeitlich angenommenen Lösungen, wenn eine zeitvariante Datenstatistik vorliegt und sind daher sehr ungenau.
- 5 4. Das "learning of learning" Verfahren ist zwar adaptiv und kann effektiv mit Nichtstationarität umgehen, es ist aber nur für den Fall einer diskreten Fehlerfunktion definiert und ist daher in dieser Form für die Praxis nicht einsetzbar.
- 10 Der Erfindung liegt demgemäß die Augabe zugrunde, ein Verfahren der eingangs genannten Gattung anzugeben, welches sich on-line an Instationaritäten in der Datenstatistik für die Anwendungen Regression, Klassifikation und Quellentrennung adaptiert.
- 15 Diese Aufgabe wird mit den im Anspruch 1 angegebenen Maßnahmen gelöst.
- Die "learning of learning" Idee wurde erfindungsgemäß erweitert und funktioniert nun auch für den Fall, daß (a) eine kontinuierliche Fehlerfunktion und/oder (b) die Hessematrix gegeben ist oder weder (a) noch (b) gegeben ist und entweder (c) ein Fluß definiert werden kann, oder (d) daß keine weitere Information vorliegt, und das entmischte Signal direkt zur Berechnung von Kovarianzmatrizen verwendet wird. Das bedeutet, daß das erfindungsgemäß Verfahren ein 20 allgemeiner On-line-Algorithmus ist, der sich selbstständig an Änderungen der Datenstatistik anpassen kann. Es kann außerdem gezeigt werden, daß sich der erweiterte Algorithmus nach dem erfindungsgemäß Verfahren in den Fällen (a) bis 25 (d) beweisbar optimal schnell an veränderte Bedingungen an-

paßt: Der erfindungsgemäße On-line-Algorithmus kann zur Regression, Klassifikation und Quellentrennung eingesetzt werden. Die Konvergenzrate ist sowohl bei stationärer als auch bei instationärer Datenstatistik beweisbar optimal. Es 5 lassen sich damit auch bekannte Verfahren zur Quellentrennung als On-line-Lernproblem formulieren und durch das erfindungsgemäße Verfahren mit optimaler Konvergenzrate instationäre Mischungen entmischen.

Das erfindungsgemäße Verfahren bzw. die Vorrichtung zur 10 Durchführung des Verfahrens weist dabei insbesondere die folgenden Vorteile auf:

- Es ist in der Lage, unüberwacht - also ausschließlich datengetrieben - zu lernen und selbständig, schnell und präzise Instationaritäten in der Datenstatistik zu erkennen und zu verfolgen. Dabei benötigt es im Extremfall nur einen Datenpunkt pro Zeit zur Verarbeitung.
- Die Schätzung der Statistik ist in Echtzeit möglich.
- Weiterhin gewährleistet es auch für stationäre Datenstatistiken eine optimale Konvergenz der Schätzung. Der zugrundeliegende On-line-Algorithmus ist einfach implementierbar, und hat beweisbar optimales Konvergenzverhalten für alle Fälle stationärer und instationärer Datenstatistik (asymptotisches Erreichen der 25 Cramer-Rao Schranke).
- Es erreicht eine Adaptivität mit optimaler Adoptionsgeschwindigkeit.

- 10 -

- Jeder Quellentrennungsalgorithmus, der als On-line-Lernproblem formuliert werden kann, kann in das erfundungsgemäße On-line-Verfahren umformuliert werden und ist dann auch in der Lage, nichtstationäre Mischungen
5 zu erkennen und zu verfolgen.
- Der On-line-Algorithmus arbeitet in den Fällen, wenn weder (a) noch (b) gegeben ist und nur (c) ein Fluß definiert werden kann. Selbst wenn keine weitere Information vorliegt, arbeitet das erfundungsgemäße Verfahren zuverlässig, indem (d) das entmischte Signal zur Berechnung von Kovarianzmatrizen verwendet wird.
10
- Der On-line-Algorithmus ist auch für Regression und Klassifikation einfach anwendbar.
- Das Verfahren ist in Hardware einfach und kostengünstig realisierbar.
15

Andere vorteilhafte Weiterbildungen der Erfindung sind in den Unteransprüchen gekennzeichnet bzw. werden nachstehend zusammen mit der Beschreibung der bevorzugten Ausführung der Erfindung anhand der Figur näher dargestellt. Es zei-
20 gen:

Figur 1 ein Ausführungsbeispiel einer Vorrichtung zur Durchführung des Verfahrens in Blockschaltung,

Figur 1a ein Detail der in Figur 1 dargestellten Vorrich-
tung,

25 Figuren 2 und 3 Signaldarstellungen zu dem in den Figuren 1 und 1a dargestellten Ausführungsbeispiel,

- 11 -

Figuren 4 bis 10 weitere Signaldarstellungen zu dem erfindungsgemäßen Verfahren.

Im folgenden werden zunächst die technischen Einzelheiten des erfindungsgemäßen Verfahrens erläutert:

- 5 Ausgegangen wird von einer Reihe von unabhängigen Beispielen in Form von Signalwerten $(x_1, y_1), (x_2, y_2), \dots$ als Eingangsgrößen. Ziel des Lernprozesses ist es dabei, eine Schätzung für den Parameter \hat{w} zu erhalten, welche die in den Daten inhärente Regel möglichst gut simulieren kann. Daher modifiziert das Verfahren seinen Parameter \hat{w} , zur Zeit t in \hat{w}_{t+1} , indem nur das nächste von der Statistik gegebene Beispiel (x_{t+1}, y_{t+1}) verwandt wird.

Dazu wird eine differenzierbare Fehlerfunktion $l(x, y; w)$ verwendet, die die Performanz des Schätzers mit Parameter w bewertet. Die Größe

$$R(w) = \langle l(x, y; w) \rangle \quad (5)$$

ist der "expected loss" oder der Generalisierungsfehler des Schätzers, das durch den Parameter w definiert ist und $\langle \rangle$ beschreibt die Mitteilung über die Verteilung der Beispiele (x, y) . Der Parameter w^* des besten Schätzers ist gegeben durch

$$w^* = \arg \min R(w). \quad (6)$$

Es wird der folgende stochastische Gradientenabstieg

- 12 -

$$\hat{w}_{t+1} = \hat{w}_t - \eta_t C(\hat{w}_t) \frac{\partial}{\partial w} l(x_{t+1}, y_{t+1}; \hat{w}_t) \quad (7)$$

benutzt, wobei η_t die Lernrate ist, die auch von der Zeit t abhängig sein kann und $C(\hat{w}_t)$ eine positiv-definite Matrix ist, die auch von \hat{w}_t abhängen kann. Die Matrix C spielt 5 die Rolle eines Riemannschen metrischen Tensors im Raum der $\{w\}$.

Wenn η_t gleich einer kleinen Konstante ist, dann konvergiert $E[\hat{w}_t]$ nach w^* und die Varianz $\text{Var}[\hat{w}_t]$ nähert sich einer nicht verschwindenden Matrix der Ordnung $O(\eta)$. Das 10 heißt, \hat{w}_t fluktuiert um w^* . Wenn $\eta_t=c/t$ (abgekühlte Lernrate), dann konvergiert \hat{w}_t lokal nach w^* . Wenn sich wie oben beschrieben die Lernregel ändert, dann kann die wie $\eta_t=c/t$ gewählte Lernrate jedoch dem Wechsel der Lernregel nicht folgen, da $\eta_t=c/t$ zu klein wird und in der Praxis die ver- 15 änderte optimale Lösung nicht mehr erreichen kann.

Die Idee einer adaptiven Lernrate η_t wurde "learning of the learning rule" genannt (Sompolinsky et al. 1995). Dieses Konzept wird bei dem erfindungsgemäßen Verfahren auf beliebige differenzierbare Fehlerfunktionen verallgemeinert. Da- 20 zu wird folgender Algorithmus zugrundengelegt:

$$\hat{w}_{t+1} = \hat{w}_t - \eta_t K^{-1}(\hat{w}_t) \frac{\partial}{\partial w} l(x_{t+1}, y_{t+1}; \hat{w}_t) \quad (8)$$

$$\eta_{t+1} = \eta_t + \alpha \eta_t \left(\beta \left(l(x_{t+1}, y_{t+1}; \hat{w}_t) - \hat{R} \right) - \eta_t \right) \quad (9)$$

wobei α und β Konstanten sind, $K(\hat{w}_t)$ ist die Hessematrix des Erwartungswertes der Fehlerfunktion $\partial^2 R(\hat{w}_t)/\partial w \partial w$ und \hat{R} ist ein Schätzer von $R(w^*)$. Wenn der Fehler explizit als Funktion von x und y gegeben ist, kann \hat{R} rekursiv als 5 Schätzwert angenommen werden durch

$$\hat{R}_{t+1} = (1-\gamma)\hat{R}_t + \eta l(x_{t+1}; y_{t+1}; \hat{w}_t) \quad (10)$$

Intuitiv wird damit die Lernrate η aus Gl. (9) durch den verbleibenden Fehler kontrolliert. Wenn der Fehler groß ist, dann nimmt η einen großen Wert an. Wenn er klein ist, 10 d.h. der geschätzte Parameter nahe beim Optimum ist, dann geht η automatisch gegen 0.

Um das dynamische Verhalten von (\hat{w}_t, η_t) zu erhalten, wird eine kontinuierliche Version der obigen Gleichungen benutzt, über die bezüglich des Input-output-Paars (x_t, y_t) 15 mittelt wurde

$$\frac{d}{dt} w_t = -\eta_t K(w_t)^{-1} \left\langle \frac{\partial}{\partial w} l(x, y; w_t) \right\rangle, \quad (11)$$

$$\frac{d}{dt} \eta_t = \alpha \eta_t \left(\beta \left\langle l(x, y; w_t) - \hat{R} \right\rangle - \eta_t \right). \quad (12)$$

Es gilt $\left\langle \partial l(x, y; w^*) / \partial w \right\rangle = 0$, und so ergibt sich asymptotisch

$$\left\langle \frac{\partial}{\partial w} l(x, y; w_t) \right\rangle \cong K(w_t - w^*),$$

$$\langle l(x, y; w_t) - \hat{R} \rangle \equiv R(w^*) - \hat{R} + \frac{1}{2} (w_t - w^*)^T K^* (w_t - w^*),$$

wobei $K^* = \partial^2 R(w^*) / \partial w \partial w$. Unter der Annahme, daß $R(w^*) - \hat{R}$ klein ist und $K(w_t) \equiv K^*$, wird erhalten

$$\frac{d}{dt} w_t = -\eta_t (w_t - w^*) \quad (13)$$

$$5 \quad \frac{d}{dt} \eta_t = \alpha \eta_t \left(\frac{\beta}{2} (w_t - w^*)^T K^* (w_t - w^*) - \eta_t \right) \quad (14)$$

Es wird eine quadratische Fehlergröße eingeführt

$$e_t = \frac{1}{2} (w_t - w^*)^T K^* (w_t - w^*), \quad (15)$$

wobei sich aus Gl. (13) und Gl. (14) ergibt

$$\dot{e}_t = -2\eta_t e_t, \quad \dot{\eta}_t = \alpha \beta \eta_t e_t - \alpha \eta_t^2. \quad (16)$$

10 Das Verhalten des Algorithmus als Gleichungssystem (16) lässt sich wie folgt deuten: Der Ursprung $(0,0)$ ist ein Attraktor seiner Dynamik und das Attraktionsbassin hat eine fraktale Grenze. Wird von einem adäquaten Anfangswert gestartet, hat das Gleichungssystem die Form

$$15 \quad e_t = \frac{1}{\beta} \left(\frac{1}{2} - \frac{1}{\alpha} \right) \cdot \frac{1}{t} \quad (\alpha > 2), \quad \text{und} \quad \eta_t = \frac{1}{2} \cdot \frac{1}{t}. \quad (17)$$

Hierbei ergeben sich als bedeutsame Vorteile:

- Die $1/t$ Konvergenzrate des Verallgemeinerungsfehlers e_t ist die optimale (schnellstmögliche) Ordnung der

- 15 -

Konvergenz jedes beliebigen Schätzers \hat{w}_t , der zum Optimum w^* konvergiert.

- Gl. (13) und (14) geben einen allgemeinen On-line-Algorithmus der schnellstmöglichen Konvergenzrate.
- 5 - Die Lösung beruht darauf, den skalaren Abstand e_t im Gewichtsraum zu verwenden.
- Damit der obige Algorithmus funktionieren kann, müssen die Größen K^* , l verfügbar sein.

Insbesondere der letzte Punkt ist in der Praxis oft nicht 10 erfüllt, so daß der gefundene Algorithmus weiter verallgemeinert werden muß.

Vielfach sind in der Praxis weder die Hessematrix K^* noch der Erwartungswert des Fehlers noch das Minimum des Erwartungswerts des Fehlers bekannt. Weiterhin gibt es in 15 einigen Anwendungen (z.B. Quellentrennung) noch nicht einmal eine explizite Fehlerfunktion. Ein verallgemeinerter Lernalgorithmus ist daher der Folgende

$$\hat{w}_{t+1} = \hat{w}_t - \eta_t f(x_{t+1}, y_{t+1}; \hat{w}_t), \quad (18)$$

wobei der Fluß f die Modifikation der Parameter bestimmt, 20 wenn ein neues Beispiel (x_{t+1}, y_{t+1}) gegeben ist. Hierbei muß nicht die Existenz einer Fehlerfunktion angenommen werden, sondern es wird nur ein gemittelter Fluß benötigt, der am Optimum verschwindet, i.e.

$$\langle f(x, y; w^*) \rangle = 0$$

Für den Fall, daß eine Fehlerfunktion gegeben ist, kann der Gradient der Fehlerfunktion als Fluß eingesetzt werden. Es wird nun die gemittelte, kontinuierliche Version von Gleichung (18) benutzt und um das Optimum

$$5 \quad \frac{d}{dt} w_t = -\eta_t \langle f(x, y; w_t) \rangle \cong -\eta_t K^* (w_t - w^*) \quad (19)$$

entwickelt, wobei $K^* = \langle \partial f(x, y; w^*) / \partial w \rangle$ ist. Unter der Voraussetzung, daß ein Vektor v mit der Eigenschaft

$$v^T K^* = \lambda v^T \quad (20)$$

und weiterhin

$$10 \quad \xi_t = \langle v^T f(x, y; w_t) \rangle \cong v^T K^* (w_t - w^*) \quad (21)$$

definiert ist, kann die Dynamik der ξ als

$$\frac{d}{dt} \xi_t = -\lambda \eta_t \xi_t, \quad (22)$$

approximiert werden. Mit den Variablen ξ wird eine Modifikation der Lernregel für η wie folgt festgelegt

$$15 \quad \eta_{t+1} = \eta_t + \alpha \eta_t (\beta |\xi_t| - \eta_t), \quad (23)$$

mit einer kontinuierlichen Version als

$$\frac{d}{dt} \eta_t = \alpha \eta_t (\beta |\xi_t| - \eta_t), \quad (24)$$

ξ läßt sich dabei als eine eindimensionale Pseudodistanz verstehen, bei der der mittlere Fluß f auf eine einzelne

Richtung v projiziert wird. Es wurde nun eine "intelligente Projektion" so gefunden, bei der es hinreichend ist, alle Dynamik des Flusses entlang dieser Richtung zu beobachten. Die Variable ξ ist entweder immer positiv oder immer negativ, abhängig davon, welcher Anfangswert gewählt wird; η ist positiv. Die beiden Gleichungen (22) und 24) sind damit äquivalent zum Gleichungssystem (16). Somit sind ihre asymptotischen Lösungen gegeben als

$$\xi_t = \frac{1}{\beta} \left(\frac{1}{\lambda} - \frac{1}{\alpha} \right) \cdot \frac{1}{t}, \quad \eta_t = \frac{1}{\lambda} \cdot \frac{1}{t}. \quad (25)$$

10 Wiederum ähnlich wie im letzten Abschnitt wurde gezeigt, daß der Algorithmus konvergiert, jedoch diesmal ohne eine explizite Fehlerfunktion oder eine Hessematrix zu benutzen.

In diesem Algorithmus ist ein wichtiges Problem, eine geeignete Projektion zu finden. Dafür wird von folgenden Annahmen ausgegangen: (1) der minimale Eigenwert der Matrix K^* ist genügend viel kleiner als der zweitgrößte Eigenwert und (2) daher wird sich der Parametervektor \hat{w}_t nach einer großen Anzahl Iterationen von der Richtung des minimalen Eigenvektors von K^* dem Optimum nähern. Unter diesen Bedingungen kann die Entwicklung des geschätzten Parameters als eindimensionaler Prozess verstanden werden und es können beliebige Vektoren als v verwendet werden, außer den zum minimalen Eigenvektor orthogonalen Vektoren. Der beste und effizienteste Vektor wird natürlich der minimale Eigenvektor selbst sein, der (für eine große Anzahl von Iterationen) durch

$$v = \langle f \rangle / \| \langle f \rangle \|.$$

approximiert werden kann. Daher können wir

$$\xi = \|\langle f \rangle\| \quad (26)$$

einsetzen. Ersetzt man nun den instantanen Mittelwert des Flusses durch einen gleitenden (leaky) Mittelwert, dann erhält man folgenden modifizierten Algorithmus

$$\hat{w}_{t+1} = \hat{w}_t - \eta_t f(x_{t+1}, y_{t+1}; \hat{w}_t), \quad (27)$$

$$r_{t+1} = (1 - \delta)r_t + \delta f(x_{t+1}, y_{t+1}; \hat{w}_t), \quad (0 < \delta < 1) \quad (28)$$

$$\eta_{t+1} = \eta_t + \alpha \eta_t (\beta \|r_{t+1}\| - \eta_t), \quad (29)$$

wobei δ den Grad des gleitenden (leaky) Mittelwertes kontrolliert und r als Hilfsvariable verwandt wird, um den gleitenden (leaky) Mittelwert des Flusses zu berechnen. Dieser Satz Lernregeln konvergiert auch asymptotisch und ist sehr einfach zu berechnen. Weiterhin induziert der gleitende (leaky) Mittelwert Fluktuationen, die sich in praktischen Anwendungen als sehr hilfreich erweisen.

- In den Fällen, in denen auch kein Fluß angenommen werden soll, arbeitet der On-line Algorithmus ebenfalls, indem das entmischte Signal zur Berechnung von Kovarianzmatrizen herangezogen wird.
- Die zeitversetzte Kovarianzmatrix eines stationären, zeitdiskreten stochastischen Prozesses ist definiert durch den Erwartungswert des äuéreren Produktes der Datenmatrix mit sich selbst.

$$K_t^r(\bar{x}) = E\{\bar{x}_t \bar{x}_{t+r}^T\} \quad (30)$$

Die Zeitversetzung τ kann dabei auch den Wert Null annehmen. Auf der Diagonalen der Kovarianzmatrix stehen die Werte der Autokorrelationsfunktion zum Zeitpunkt t . Die Nebendiagonalelemente bilden die Werte der Kreuzkorrelationsfunktion der jeweiligen Signale zum Zeitpunkt t . Daraus folgt eine wichtige Eigenschaft der Kovarianzmatrix, sie ist eine symmetrische Matrix.

5 Im Fall eines nichtstationären Prozesses kann die Kovarianzmatrix adaptiv online geschätzt werden

$$10 \quad K_{t+1}^r(\bar{x}) = (1 - \frac{1}{\eta_t^r}) K_t^r(\bar{x}) + \frac{1}{\eta_t^r} \bar{x}_{t+1} \bar{x}_{t+1}^T \quad (31)$$

$$\eta_{t+1}^r = \eta_t^r + \alpha \eta_t^r (\beta \|K_{t+1}^r\| - \eta_t^r) . \quad (32)$$

Nun kann der von Schuster und Molgedey (1994) vorgeschlagene Ansatz, das Quellentrennungsproblem mittels simultaner Diagonalisierung zweier oder mehrerer (zeitverschobener) 15 Kovarianzmatrizen zu lösen, zu einem einfachen und effizienten online Verfahren erweitert werden.

Dazu wird die nach Gleichung (31) und (32) bestimmte Folge von Kovarianzmatrizen mit einem online Verfahren simultan diagonalisiert: (1) In einem online whitening Schritt und 20 (2) in einer online Berechnung einer orthogonalen Transformation durch die adaptive (online) Lösung eines Eigenwertproblems (zum Beispiel mit Sanger's Rule (T.D. Sanger, in *Neural Networks*, Bd. 2, S. 459 (1989)), oder mit einem anderen numerischen adaptiven Eigenwertlösungsverfahren) ge- 25 mäß

$$W_{t+1} = W_t - \eta_t^r (W_t K_{t+1}^r W_t^T - 1) \quad (33)$$

- 20 -

$$C_{t+1} = W_{t+1} K_{t+1}^{r_2} W_{t+1}^T \quad (34)$$

$$V_{t+1} = V_t + \eta_{t+1}^{r_2} [C_{t+1} V_t - V_t - \text{triu}(V_t^T C_{t+1} V_t)] \quad (35)$$

$$\hat{W}_{t+1} = V_{t+1} W_{t+1} \quad (36)$$

Dabei ist triu eine Operation, die die obere Dreiecksmatrix
5 auswählt, und die Elemente unterhalb der Diagonalen zu Null
setzt.

Zusammenfassen besteht die letztgenannte Variante daraus

1. adaptiv die Kovarianzmatrix und mindestens eine zeitver-
schobene Kovarianzmatrix zu schätzen,
- 10 2. die mindestens zwei Kovarianzmatrizen simultan zu dia-
nalisieren, so daß eine Entmischungsmatrix gewonnen
wird, die die Signalquellen trennt.

Die Anwendung als Verfahren soll nachfolgend anhand des in
Figur 1 wiedergegebenen Blockschaltbilds näher erläutert
15 werden. Hierbei soll die folgende beispielhafte Illustrati-
on eines Ausführungsbeispiels dieser Lösung in der akusti-
schen Quellentrennung zugrunde gelegt werden:

Zwei Sprecher bewegen sich im Raum und unterhalten sich da-
bei. Über zwei Mikrofone werden ihre (instationär) gemisch-
20 ten Sprachsignale aufgenommen. Diese werden der nachfolgend
dargestellten Verarbeitungseinheit zugeführt. Trotz der
ständigen veränderten Mischung, gelingt es, die akustischen
Einzelsprechersignale aus der Mischung herauszufiltern. Am
Ausgang erscheinen die getrennten Sprachsignale der Ein-
25 zelsprecher als entmischte Signale.

(Andere Anwendungen - insbesondere aus dem Medizinbereich (EEG, MEG) - werden weiter unten erläutert.

Eine Vorrichtung zur Durchführung des erfindungsgemäßen Verfahrens beruht dabei auf einem System, welches einen Lösungsalgorithmus für die angegebenen Gleichungen beinhaltet. Die Realisierung der konkreten Schaltung kann in paralleler, serieller oder auch gemischt parallel/serieller Anordnung erfolgen. Die Elemente der Hardware sind als arithmetische Baugruppen zur Ausführung der genannten Algorithmen aufgebaut. Es handelt sich dabei um die Gleichungen (27) bis (29) sowie (8) bis (10), und (31) bis (36). Dazu lassen sich entweder komplett Prozessoren mit den entsprechenden, in Speichern festgehaltenen Programmen verwenden oder aber spezielle Hardwarekomponenten, welche Addierer, Multiplizierer und ein Modul zur Inversion der Matrix $(1+T_i)^{-1}$ aufweisen. Weiterhin muß eine Vorrichtung zur Berechnung der Norm des Flusses vorgesehen sein.

Die Funktionsweise ist dabei wie folgt:

Die "learning of learning" Idee von Sompolinski et al. 1995 wird funktionell für den allgemeinen Fall erweitert, daß (a) eine kontinuierliche Fehlerfunktion und/oder (b) die Hessematrix gegeben ist oder weder (a) noch (b) gegeben ist und nur (c) ein Fluß definiert werden kann, oder (d) Kovarianzmatrizen berechnet werden.

Hierfür sind folgende Bedingungen maßgeblich:

- i. Sind (a) kontinuierliche Fehlerfunktion und (b) die Hessematrix gegeben, so werden die Gleichungen (8) bis (10) verwendet.

ii. Ist nur eine kontinuierliche Fehlerfunktion ohne Hessematrix gegeben, so werden die Gleichungen (27) bis (29) verwendet, wobei der Gradient der Fehlerfunktion als Fluß dient und $\xi = \|\langle f \rangle\|$.

5 iii. Ist nur (c) ein Fluß definiert, so wird Gleichungen (27) bis (29) verwendet und $\xi = \|\langle f \rangle\|$.

iv. Ist auch kein Fluß definiert, so werden die Gleichungen (31) bis (36) verwendet.

Der Fluß kann dabei beliebig, entsprechend den Ansätzen von 10 Bell et al. (1995), Jutten et al. (1991), Molgedey & Schuster 1994 oder beliebigen anderen Quellentrennungsalgorithmen gewählt werden.

Dabei bildet die Projektion auf $\xi = \|\langle f \rangle\|$ ein sehr wichtiges Element des erfindungsgemäßen Verfahrens. Für die genannten 15 Fälle kann damit bewiesen werden, daß das erfindungsgemäße Verfahren ein optimales Konvergenzverhalten zeigt.

Bei dem in Figur 1 dargestellten Ausführungsbeispiel wird von einem Mischsignal 1 ausgegangen. Dieses Signal enthält die Signale einzelner Signalquellen gegebenenfalls mit einem nicht zu trennenden Hintergrundsignal. Aus diesem Signal werden mittels zweier Mikrofone Signalanteile als Beispieldsignale 1.1 und 1.2 entnommen. Die Möglichkeit der Entnahme weiterer Signale ist durch gestrichelte Linien angedeutet.

25 Die Signale 1.1 und 1.2 (und gegebenenfalls weitere) werden einer Entmischungsstufe 2 zugeführt, in der die Signaltrennung nach einem bekannten Algorithmus erfolgt. Die Entmi-

schungsstufe enthält bei der bevorzugten Ausführung eine Schaltmatrix, mittels der einzelne Eingänge zeitweise mit einzelnen Ausgängen unter entsprechender Gewichtung verbunden werden können. Dies kann auch mit multiplikativer oder 5 additiver Überlagerung erfolgen. In Betracht kommt auch eine reine Schaltmatrix bei der die Signalintensitäten am Ausgang durch Modulation der Impulsdauer und entsprechende Integration erzeugt wird. Die Ansteuerung der (in der Zeichnung nicht näher dargestellten) Schaltmatrix erfolgt 10 mit entsprechenden Steuersignalen, die vom Ausgang eines später zu beschreibenden entsprechenden Steuerbaustein ausgehen.

Die getrennten Ausgangssignale 3.1 und 3.2 bilden entmischtte Signale und stellen im dargestellten Beispielsfall 15 die akustischen Signale der einzelnen Klangquellen dar. Konkret sind dies hier die Sprachsignale der beiden Sprecher.

Die Beispielsignale 1.1 und 1.2 aus dem Mischsignal und die entmischten Signalen 3.1 und 3.2, sowie die im vorherigen 20 Durchgang in einer dritten Baugruppe 7 ermittelte Lernrate werden einer ersten Baugruppe zur Bewertung des entmischten Signals 4 zugeführt. In der ersten Baugruppe 4 werden in Abhängigkeit von einem aus einem Entscheidungs-Modul 5 abgeleiteten Signal entweder die Algorithmen (10) oder (28), 25 oder (31) angewendet.

Eine Fallunterscheidung bezüglich der Anwendung der verschiedenen Algorithmen wird in dem Entscheidungsmodul 5 getroffen, in der durch UND-Gatter die nachfolgend beschriebenen logischen Signalverknüpfungen realisiert sind:

- i. Sind (a) kontinuierliche Fehlerfunktion und (b) die Hessematrix gegeben, so wird ein Steuersignal ausgegeben, daß die Verwendung der Gleichungen (8) bis (10) veranlaßt.
- 5 ii. Ist nur die kontinuierliche Fehlerfunktion ohne Hessematrix gegeben, so wird ein Steuersignal ausgegeben, daß die Verwendung der Gleichungen (27) bis (29) veranlaßt, wobei der Gradient der Fehlerfunktion als Fluß dient und $\xi = \|f\|$ gilt.
- 10 iii. Ist nur (c) ein Fluß definiert, so wird ein Steuersignal ausgegeben, daß die Verwendung der Gleichungen (27) bis (29) veranlaßt.
- iv. Ist (d) weder eine kontinuierliche Fehlerfunktion noch ein Fluß definiert, so wird ein Steuersignal 15 ausgegeben, daß die Verwendung der Gleichungen (31) bis (36) veranlaßt, und Kovarianzmatrizen werden zur Bewertung des entmischten Signals verwendet.

In Figur 1a sind Details des Entscheidungs-Moduls 5 wieder-gegeben. In den Blöcken 5.1 bis 5.3 und 5.6 werden dabei 20 jeweils die Eingangsbedingungen (a) bis (d) entsprechend der vorangehenden Aufstellung ermittelt. In der Stufe 5.4 erfolgen die logischen Entscheidungen entsprechend dem vor-stehenden Schema. Die dabei notwendigen Negationen ergeben sich über die entsprechenden invertierenden Eingänge der 25 dargestellten UND-Gatter. In der Stufe 5.5 werden die Aus-gangssignale der Stufe 5.4 zu einem einzigen Signal zusam-mengefaßt, dessen Wert bestimmt, welcher der oben genannten Fälle (i) bis (iv) vorliegt.

Das Ergebnis der Bewertung des entmischten Signals sowie das Steuersignal gelangt dann an eine zweite Baugruppe 6, in der die Adaptation der Schaltmatrix erfolgt. Je nach Steuersignal werden in der Baugruppe 6 die Gleichungen (8), 5 oder (27), oder (33) bis (36) angewandt.

Diese Baugruppe steht in Wechselwirkung mit der dritten Baugruppe für die Lernratensteuerung 7. Es ist ersichtlich, daß mit den Algorithmen des erfindungsgemäßen Verfahrens sich die Lernrate, d.h. die Änderung des die Entmischungs- 10 stufe über einen Steuerbaustein 8 beeinflussenden Algorith- mus entsprechend der aktuellen Änderung der Lernrate des Systems anpaßt. Der Steuerbaustein 8 enthält die zur An- steuerung der Entmischungsstufe notwendigen elektrischen Bauelemente.

15 Die Lernrate wird also ebenfalls vom adaptiven Algorithmus geregelt, und zwar, abhängig von dem im Entscheidungsmodul 5 erzeugten Steuersignal, nach den Algorithmen (9) oder (27) oder (32). Damit kann das System Änderungen der Quel- 20 len sowohl in ihrer Intensität als auch bezüglich der Ände- rung der Intensität und anderer Signaleigenschaften optimal schnell folgen. Verläuft der Prozeß stationär, "kühlt sich die Lernrate ab".

Als Ergebnis der Beispielanwendung im Bereich akustischer Quellentrennung nichtstationärer Mischungen wurden die 25 nachfolgend beschriebenen weiteren Ergebnisse erhalten.

Als Beispiel wurden zwei Audiosignale bestehend aus Sprache und Musik (mit einer Samplingrate von 8 kHz) verwendet: "Fischers Fritz" (s_1^1) und "Donauwalzer" s_1^2 (siehe Fig. 2).

Als Signal 1 wurde der bekannte Zungenbrecher "Fischers

Fritz fischt frische Fische ..." aufgenommen, während Signal 2 ein Teil des "Donauwalzer" von J. Strauß ist. Beide akustische Quellen wurden auf dem Computer gemischt entsprechend

$$5 \quad \bar{I}_t = (1+A)\bar{s}_t \quad f. \quad 0s < t < 2,5s \quad \text{und} \quad 6,25s < t < 10s \quad (30)$$

$$\bar{I}_t = (1+B)\bar{s}_t \quad f. \quad 2,5s \leq t < 6,25 \quad (31)$$

gemischt mittels

$$A = \begin{pmatrix} 0 & 0,9 \\ 0,7 & 0 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 0 & 0,8 \\ 0,6 & 0 \end{pmatrix}$$

als Mischungsmatrizen. Das heißt, die den Daten unterliegende Statistik schaltet zweimal in diesem Datensatz. Das Ziel ist nun, die Quellen \bar{s}_t zu erhalten, indem \hat{A} und \hat{B} zunächst geschätzt werden. Dazu stehen nur die gemessenen Mischsignale \bar{I}_t zur Verfügung - alles andere ist unbekannt. Daher wird dieses Problem auch oft "blinde Quellentrennung" genannt. Eine Änderung der Mischung findet man oft in der Praxis, wenn beispielsweise ein Sprecher seinen Kopf während des Sprechens bewegt oder seine Position wechselt.

Der erfindungsgemäße adaptive On-line-Algorithmus ist besonders für dieses instationäre Szenario geeignet, da seine Adaptionsschnelligkeit nicht durch die bereits oben diskutierten generischen Nachteile einer konstanten Lernrate limitiert ist - wie in den Ansätzen von Bell et al. 1995, Jutten et al. 1991, Molgedey & Schuster 1994. Es sei \bar{u}_t der Vektor des entmischten Signals

$$\bar{u}_t = (1 + T_t)^{-1} \bar{I}_t \quad (32)$$

wobei T die geschätzte Mischungsmatrix ist. Basierend auf dem Algorithmus von Molgedey & Schuster 1994 wird als Lernregel für T_t

5
$$\Delta T_t^j \propto f(\langle I_t^j u_t^j \rangle, \langle u_t^i u_t^j \rangle, \langle I_t^j u_{t-1}^j \rangle, \langle u_t^i u_{t-1}^j \rangle, \dots)$$

$$\propto \langle I_t^j u_t^j \rangle \langle u_t^i u_t^j \rangle + \langle I_t^j u_{t-1}^j \rangle \langle u_t^i u_{t-1}^j \rangle, \quad i, j = 1, 2, \quad i \neq j, \quad (33)$$

benutzt, wobei die instantanen Mittelwerte durch die oben definierten gleitenden (leaky) Mittelwerte ersetzt werden.

$$\langle I_t^j u_t^j \rangle_{\text{leaky}} = (1 - \epsilon) \langle I_{t-1}^j u_{t-1}^j \rangle_{\text{leaky}} + \epsilon I_t^j u_t^j.$$

10 Hier ist zu bemerken, daß die notwendigen Bestandteile des Flusses f aus Gleichungen (27) bis (28) für diesen Fall einfach die gleichzeitigen oder zeitversetzten Korrelationen sind und η_t gemäß Gleichung (29) berechnet wird.

Fig. 3 zeigt das Resultat der Simulation (Details der Parameterwahl sind in der Bildunterschrift erklärt). Nach einer kurzen Zeit ($t = 0,4$ s) von großer Lernrate η und starken Fluktuationen in η wurde die Mischungsmatrix korrekt geschätzt. Bis $t = 2,5$ s wird die Lernrate automatisch entsprechend $1/t$ (siehe untere Darstellung in Fig. 3) 15 "abgekühlt", wie auch in Gl.(25) vorhergesagt, d.h. es wird auch im Experiment die optimale "Abkühlrate" gefunden. Am Schaltpunkt, an dem simples "Abkühlen" versagt hätte, vergrößert der erfindungsgemäße adaptive Algorithmus, wie gewünscht, die Lernrate drastisch und ist damit auch in der 20 25 Lage, den durch das Schalten veränderten Mischungen inner-

halb von 0,1 s (bzw. 0,2 s beim zweiten Schalten) zu folgen. Danach wird die Lernrate, wie gewünscht, wiederum "abgekühlt".

Vergleicht man Misch-, Original- und entmischt Signal in 5 Fig. 2, dann bestätigt sich die schnelle und genaue Schätzung, die bereits für die Mischungsmatrix beobachtet werden konnte. Auch ein Anhören bestätigt die Effizienz des Verfahrens, für Sekundenbruchteile sind beide Signale hörbar, danach nur noch das entmischt Signal. Nach dem Schalten 10 sind wiederum einen kurzen Moment lang beide Signale hörbar. Der Entmischungsprozess geht dabei so schnell vor sich, daß der Hörer das Gefühl hat, ein Signal würde einfach vom Quellentrennungsalgorithmus "ausgeschaltet". Erfolgt eine Adaption in der Größenordnung von 0,1 s, dann 15 wird dies als Knacken empfunden. Der Rauschabstand im entmischten Signal beträgt für das Musiksignal 36 dB und für das Sprachsignal 28 dB.

Das Verfahren wurde natürlich auch für andere Instationaritäten sowohl theoretisch als auch experimentell untersucht. 20 Die weiteren Figuren 6 bis 8 zeigen das Ergebnis weiterer Versuche, bei denen ein kontinuierlicher Übergang und einen weiteren Schaltübergang untersucht wurden.

Für das erfindungsgemäße Verfahren kann die theoretische Zeit, die der Algorithmus braucht, um einer Änderung zu 25 folgen, berechnet werden, so daß eine gute Anpassung an die praktischen Voraussetzungen erfolgen kann.

Zusammenfassend ergibt sich damit ein exzellentes Adaptionsverhalten des Algorithmus, das sich in vielen weiteren, hier nicht beschriebenen, Experimenten reproduzieren ließ.

Das Verfahren wurde im Experiment bereits erfolgreich zur Trennung von akustischen Signalen angewandt für

- die Trennung mehrerer Sprecher,
- die Trennung von Sprache und Musik,
- 5 - die Trennung von Sprache und Hintergrundgeräusch.

Anwendungen des Verfahrens sind für Hörgeräte günstig, welche damit die dynamische Änderungen der akustischen Mischverhältnisse verfolgen können und in Echtzeit durch geeignete Hardware das Hörsignal als Auswahl einzelner Quellen 10 verstärken.

Weiterhin ist das Verfahren für Anwendungen im Verkehrsmittel- (Automobil-)bereich besonders geeignet. Durch eine Online-Trennung von Sprache und Hintergrundgeräusch kann der Störpegel bei einem Telefonat reduziert werden. Bei sprachgesteuerten Eingaben im Automobil kann die Fehlerrate in 15 der Erkennungsleistung durch Reduzierung des Hintergrundgeräuschpegels herabgesetzt werden. Anwendungsmöglichkeiten bestehen weiterhin auf allen Gebieten der Nachrichtentechnik, bei denen einzelne Signalquellen voneinander oder von 20 einem Hintergrundsignal getrennt werden sollen.

Außerhalb der akustischen Quellentrennung ergeben sich Anwendungen vor allem in der Medizintechnik. Bei der Aufnahme von Multielektroden EEGs (Elektroenzephalogrammen) oder MEGs (Magnetenzephalogrammen) sind die gemessenen Signale 25 stets instationäre Überlagerungen von Gehirnaktivität. Man kann sich vorstellen, daß sich verteilte neuronale Aktivitäten- überlagern und nur so von EEG oder MEG-Geräten gemessen werden können. Da neuronale Aktivitäten hochgradig

- instationär sind, können bisherige auf stationäre Quellen-trennung gestützte Verfahren wie das von Makeig et al. 1996, die instationären Daten niemals korrekt erklären. Hier be-steht ein großes Anwendungspotential der erfindungsgemäßen
- 5 Lösung, da das vorgestellte Verfahren Änderungen der Signa-le und der Signalorte mit optimaler Schnelligkeit verfolgen kann. Es wäre denkbar, neuronale Aktivitätsmuster des Ge-hirns On-line-und in Echtzeit verfolgen und bestimmen zu können.
- 10 In modernen nichtinvasiven Untersuchungsverfahren der Hirn-aktivität werden beispielsweise SQUID-Magnetometer einge-setzt. Dies sind hochempfindliche Sensoren für magnetische Felder. Damit ist es möglich, die sehr schwachen Magnetfel-der (Feldstärken im Femtotesla Bereich), die von den elek-
15 trischen Ionenströmen der aktivierten Nervenzellen erzeugt werden, nachzuweisen. Da das Magnetfeld der Erde um einiges stärker ist, befinden sich die Sensoren in einem magnetisch abgeschirmten Raum. Dennoch kann es zur Beeinträchtigung der Messungen durch körpereigene biomagnetische Störquellen
20 kommen. So verursachen z. B. die Augenbewegungen (EOG) Ver-fälschungen der Messungen von evozierten Feldern im visuel-
len Cortex. Herzschlag (EKG) und Muskeltätigkeit (EMG) sind die Hauptstörquellen bei Messungen am peripheren Nervensystem.
- 25 Der Einsatz der adaptiven Quellentrennung ermöglicht es so-wohl stationäre als auch instationäre Störer von Augenbewe-gungen (EOG), Herzschlag (EKG) und Muskeltätigkeit (EMG) - also physiologische Störsignale - und ebenso technische Störsignale aus der Experimentieranordnung (z. B. Stimu-
30 lusimpulse bei funktionellen Reizreaktionsexperimenten) präzise zu identifizieren und zu eliminieren.

Als Ergebnis einer Anwendung im Bereich magnetische Quellentrennung wurden die nachfolgend beschriebenen Ergebnisse erhalten. Figur 9 zeigt einige SQUID Aufnahmesignale. Das gesuchte magnetische Signal, das von aktivierten Nervenzellen 5 ausgeht, ist von den oben beschriebenen Störquellen überlagert, die um Größenordnungen stärker sind. Das Stör- signal vom Herzen ist auf der Figur gut zu erkennen.

Eine Quellentrennung nach Methode (d) ermöglichte beispielsweise die Extraktion der Störsignale, die von (a) 10 Herz, (b) Stimulus und (c) Netzbrumm (inklusive (d) Harmonischer) herrühren, wie in Figur 10 gezeigt. Durch Subtraktion dieser Störsignale von der Originalmessung wird das gesuchte Signal erhalten.

In medizinischen Untersuchungen, beispielsweise in 15 Schlaflaboratorien, stehen verschiedene Signaltypen zur Verfügung (z.B. EEG, EKG, Atmung, EOG, Blutdruck usw.) Diese Signale sind instationäre Überlagerungen verschiedener unterliegender wahrer Signale, die man nicht kennt. Bisher wurde in der Datenanalyse ausschließlich von einem Statio- 20 naritätsparadigma ausgegangen. Es ist nun möglich, die Instationaritäten zu berücksichtigen.

Eine weitere Anwendung besteht in der Tele-Identifikation (Multiobjektidentifikation) Das System besteht aus einer fest montierten Basisstation und zahlreichen mobilen Minisendern. Die Basisstation sendet permanent eine elektromagnetische Welle z.B. der Frequenz 120 Hz. Kommt nun einer 25 der Minisender so nah an die Basisstation heran, daß die Energie ausreicht, den Minisender in Betrieb zu setzen, sendet dieser den auf (s)einem Speicherchip befindlichen 30 Code. Dieses Codesignal wird von der Basisstation empfangen

und decodiert. Der empfangene Code kann jetzt beispielsweise dazu benutzt werden, den Minisender (und damit das Objekt) zu identifizieren und weitere Informationen aus einer Datenbank abzurufen.

5 Der Einsatz von Verfahren zur Quellentrennung wird insbesondere dann nötig, wenn sich mehrere Sender gleichzeitig im Bereich einer Basisstation befinden und sich die Signale überlagern. In der Praxis kommt dabei meist erschwerend hinzu, daß sich die Sender bewegen und so die Mischung der
10 Signale nicht stationär ist. Für dieses Problem bietet der erfindungsgemäße Algorithmus zur adaptiven Quellentrennung die geeignete Lösung.

Das erfindungsgemäße Verfahren beschränkt sich in seiner Anwendung nicht auf die vorstehend angegebenen bevorzugten
15 Ausführungsbeispiele. Es kann auf alle Arten von Signalen angewendet werden, welche sich in einer charakteristischen, - für das Verfahren erkennbaren Weise - von anderen (Hintergrund-) Signalen unterscheiden. Dabei besteht keine Einschränkung in der Auswahl der Signale. So können bei-
20 spielsweise auch Personen- oder Wirtschaftsdaten verwendet werden. Dabei ist eine Anzahl von Varianten günstig, welche von der dargestellten Lösung auch bei grundsätzlich anders gearteten Ausführungen Gebrauch macht.

* * * * *

Ansprüche

1. Verfahren zur nichtstationären Trennung von Signalen unterschiedlicher Quellen oder Trennung mindestens eines Signals von einem Signalhintergrund, mit einer Entmischungsstufe, die eine Schaltmatrix aufweist, die eine Anzahl von Mischungssignalanteilen aufweisenden Signaleingängen unterschiedlichen Ausgängen über Bewertungselemente jeweils additiv oder multiplikativ über einen adaptiven Algorithmus steuerbar zuordnet, wobei der adaptive Algorithmus unter Nutzung einer variablen Lernrate veränderbar ange-
5 steuert wird,
10

dadurch gekennzeichnet,

daß die Lernrate kontinuierlich unter Einschluß eines Algorithmus angepaßt wird, welcher die Schritte

- 15 i) Ermittlung einer differenzierbaren Funktion oder mindestens zweier Matrizen aus dem Signal, dem getrennten Signal und der Lernrate, die das getrennte Signal bewertet bzw. bewerten,
- ii) Adaptation der Schaltmatrix auf der Basis des Ergebnisses von (i) sowie der Lernrate,
- 20 iii) Adaptation der Lernrate auf der Basis der Ergebnisse von (i) und (ii),

umfaßt,

wobei die Abfolge der Schritte (ii) und (iii) auch umgekehrt sein kann, so daß (ii) auf der Basis der Ergebnisse von (i) und (iii), und (iii) auf der Basis des Ergebnisses von (i) sowie der Schaltmatrix ausgeführt wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß der Algorithmus die Schritte

$$i) \quad r_{t+1} = (1-\delta)r_t + \delta f(x_{t+1}, y_{t+1}; \bar{w}_t), \quad (0 < \delta < 1) \quad (28)$$

$$ii) \quad \bar{w}_{t+1} = \bar{w}_t - \eta_t f(x_{t+1}, y_{t+1}; \bar{w}_t), \quad (27)$$

$$5 \quad iii) \quad \eta_{t+1} = \eta_t + \alpha \eta_t (\beta \|r_{t+1}\| - \eta_t), \quad (29)$$

umfaßt.

3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß der Algorithmus die Schritte

$$i) \quad \bar{R}_{t+1} = (1-\gamma)\bar{R}_t + \gamma l(x_{t+1}, y_{t+1}; \bar{w}_t) \quad (10)$$

$$10 \quad ii) \quad \bar{w}_{t+1} = \bar{w}_t - \eta_t K^{-1}(\bar{w}_t) \frac{\partial}{\partial w} l(x_{t+1}, y_{t+1}; \bar{w}_t) \quad (8)$$

$$iii) \quad \eta_{t+1} = \eta_t + \alpha \eta_t (\beta (l(x_{t+1}, y_{t+1}; \bar{w}_t) - \bar{R}) - \eta_t) \quad (9)$$

umfaßt.

4. Verfahren nach Anspruch 1, dadurch gekennzeichnet,

daß die ersten beiden Schritte aus

- 15 i) der Bewertung des getrennten Signals durch Berechnung von zwei oder mehr zeitverschobenen Kovarianzmatrizen und
 ii) der Adaptation der Schaltungsmatrix durch simultane Diagonalisierung der Kovarianzmatrizen bestehen.

5. Verfahren nach Anspruch 4 **dadurch gekennzeichnet**, daß die Berechnung der Kovarianzmatrizen adaptiv nach
- $$K_{t+1}^r(\hat{x}) = (1 - \frac{1}{\eta^r}) K_t^r(\hat{x}) + \frac{1}{\eta^r} \hat{x}_{t+1} \hat{x}_{t+1}^T \quad (31)$$
- erfolgt, wobei die Zeitverschiebung mindestens einmal den Wert Null annimmt.

6. Verfahren nach Anspruch 4, **dadurch gekennzeichnet**, daß die Adaptation der Lernrate gemäß
- $$\eta_{t+1}^r = \eta_t^r + \alpha \eta_t^r (\beta \|K_{t+1}^r\| - \eta_t^r) \quad (32)$$
- erfolgt.
- 10 7. Verfahren nach Anspruch 4, **dadurch gekennzeichnet**, daß die simultane Diagonalisierung aus einem Whitenizing Schritt

$$W_{t+1} = W_t - \eta_t^{r_1} (W_t K_{t+1}^{r_1} W_t^T - I) \quad (33)$$

und einer orthogonalen Transformation

$$15 \quad C_{t+1} = W_{t+1} K_{t+1}^{r_2} W_{t+1}^T \quad (34)$$

$$V_{t+1} = V_t + \eta_{t+1}^{r_2} [C_{t+1} V_t - V_t - \text{triu}(V_t^T C_{t+1} V_t)] \quad (35)$$

besteht, so daß sich die Schaltmatrix aus

$$18 \quad \hat{W}_{t+1} = V_{t+1} W_{t+1} \quad (36)$$

ergibt.

8. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, daß mindestes eines der folgenden Kriterien für die Auswahl der Ausdrücke zur algorithmischen Berechnung verwendet wird:

- 5 Sind (a) kontinuierliche Fehlerfunktion und (b) die Hessematrix ermittelbar, so werden die Gleichungen

$$\bar{w}_{t+1} = \bar{w}_t - \eta_t K^{-1}(\bar{w}_t) \frac{\partial}{\partial w} l(x_{t+1}, y_{t+1}; \bar{w}_t) \quad (8)$$

$$\eta_{t+1} = \eta_t + \alpha \eta_t (\beta(l(x_{t+1}, y_{t+1}; \bar{w}_t) - \bar{R}) - \eta_t) \quad (9)$$

angewandt.

- 10 Ist nur eine kontinuierliche Fehlerfunktion ohne Hessematrix ermittelbar, so werden die Gleichungen

$$\bar{w}_{t+1} = \bar{w}_t - \eta_t f(x_{t+1}, y_{t+1}; \bar{w}_t), \quad (27)$$

$$r_{t+1} = (1-\delta)r_t + \delta f(x_{t+1}, y_{t+1}; \bar{w}_t), \quad (0 < \delta < 1) \quad (28)$$

$$\eta_{t+1} = \eta_t + \alpha \eta_t (\beta \|r_{t+1}\| - \eta_t), \quad (29)$$

- 15 angewandt, wobei der Gradient der Fehlerfunktion als Fluß dient und $\xi = \|\langle f \rangle\|$ ist.

Ist nur (c) ein Fluß ermittelbar, so werden die algorithmischen Ausdrücke

$$\bar{w}_{t+1} = \bar{w}_t - \eta_t f(x_{t+1}, y_{t+1}; \bar{w}_t), \quad (27)$$

$$r_{t+1} = (1-\delta)r_t + \delta f(x_{t+1}, y_{t+1}; \bar{w}_t), \quad (0 < \delta < 1) \quad (28)$$

$$\eta_{t+1} = \eta_t + \alpha \eta_t (\beta \|r_{t+1}\| - \eta_t), \quad (29)$$

angewandt und $\xi = \|\langle f \rangle\|$.

Ist (d) weder eine Fehlerfunktion noch ein Fluß ermittelbar, so werden die algorithmischen Ausdrücke

- 37 -

$$K_{t+1}^r(\hat{x}) = (1 - \frac{1}{\eta_t^r}) K_t^r(\hat{x}) + \frac{1}{\eta_t^r} \hat{x}_{t+1} \hat{x}_{t+1}^T \quad (31)$$

$$\eta_{t+1}^r = \eta_t^r + \alpha \eta_t^r (\beta \|K_{t+1}^r\| - \eta_t^r) \quad (32)$$

$$W_{t+1} = W_t - \eta_t^{r_1} (W_t K_{t+1}^{r_1} W_t^T - 1) \quad (33)$$

$$C_{t+1} = W_{t+1} K_{t+1}^{r_2} W_{t+1}^T \quad (34)$$

$$5 \quad V_{t+1} = V_t + \eta_{t+1}^{r_2} [C_{t+1} V_t - V_t - \text{triu}(V_t^T C_{t+1} V_t)] \quad (35)$$

$$\hat{W}_{t+1} = V_{t+1} W_{t+1} \quad (36)$$

angewandt.

9. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, daß die Ermittlung des Flusses aufgrund von unabhängig aus dem Signalgemisch entnommenen Signalproben und den entmischten Signalen erfolgt.

10. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, daß der Fluß zur Beeinflussung des adaptiven Algorithmus dient.

15 11. Verfahren nach einem der vorhergehenden Ansprüche, **gekennzeichnet durch** die Anwendung auf nachrichtentechnische, insbesondere akustische, Signale, medizinische Signale oder Wirtschafts-Daten.

12. Vorrichtung zur Durchführung des Verfahrens nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, daß eine Rechenschaltung vorgesehen ist, welche die Ausdrücke

$$5 \quad \vec{w}_{t+1} = \vec{w}_t - \eta_t f(x_{t+1}, y_{t+1}; \vec{w}_t), \quad (27)$$

$$r_{t+1} = (1 - \delta)r_t + \delta f(x_{t+1}, y_{t+1}; \vec{w}_t), \quad (0 < \delta < 1) \quad (28)$$

$$\eta_{t+1} = \eta_t + \alpha \eta_t (\beta \|r_{t+1}\| - \eta_t), \quad (29)$$

und/oder

$$\vec{w}_{t+1} = \vec{w}_t - \eta_t K^{-1}(\vec{w}_t) \frac{\partial}{\partial w} l(x_{t+1}, y_{t+1}; \vec{w}_t) \quad (8)$$

$$10 \quad \eta_{t+1} = \eta_t + \alpha \eta_t (\beta (l(x_{t+1}, y_{t+1}; \vec{w}_t) - \vec{R}) - \eta_t) \quad (9)$$

und/oder

$$K_{t+1}^r(\vec{x}) = (1 - \frac{1}{\eta_t^r}) K_t^r(\vec{x}) + \frac{1}{\eta_t^r} \vec{x}_{t+1} \vec{x}_{t+1}^T \quad (31)$$

$$\eta_{t+1}^r = \eta_t^r + \alpha \eta_t^r (\beta \|K_{t+1}^r\| - \eta_t^r) . \quad (32)$$

$$W_{t+1} = W_t - \eta_t^{r_1} (W_t K_{t+1}^{r_1} W_t^T - 1) \quad (33)$$

$$15 \quad C_{t+1} = W_{t+1} K_{t+1}^{r_1} W_{t+1}^T \quad (34)$$

$$V_{t+1} = V_t + \eta_{t+1}^{r_2} [C_{t+1} V_t - V_t - \text{triu}(V_t^T C_{t+1} V_t)] \quad (35)$$

$$\hat{w}_{t+1} = V_{t+1} W_{t+1} \quad (36)$$

in soft- oder hardwaremäßiger Implementierung aufweist.

- 39 -

13. Vorrichtung nach Anspruch 7 oder 8, **dadurch gekennzeichnet**, daß sie ein Zusatzgerät zu oder einen Teil einer Verarbeitungseinrichtung für drahtlose oder drahtgebundene, insbesondere akustische, Signale, wie einer Anlage zur Aufbereitung von Musik, einer Hörhilfe, eines Geräusch- oder Störunterdrückungssystems und/oder eines medizinischen Gerätes und/oder eines Geräts zur Informationsaufbereitung oder -verarbeitung bildet.

* * * * *

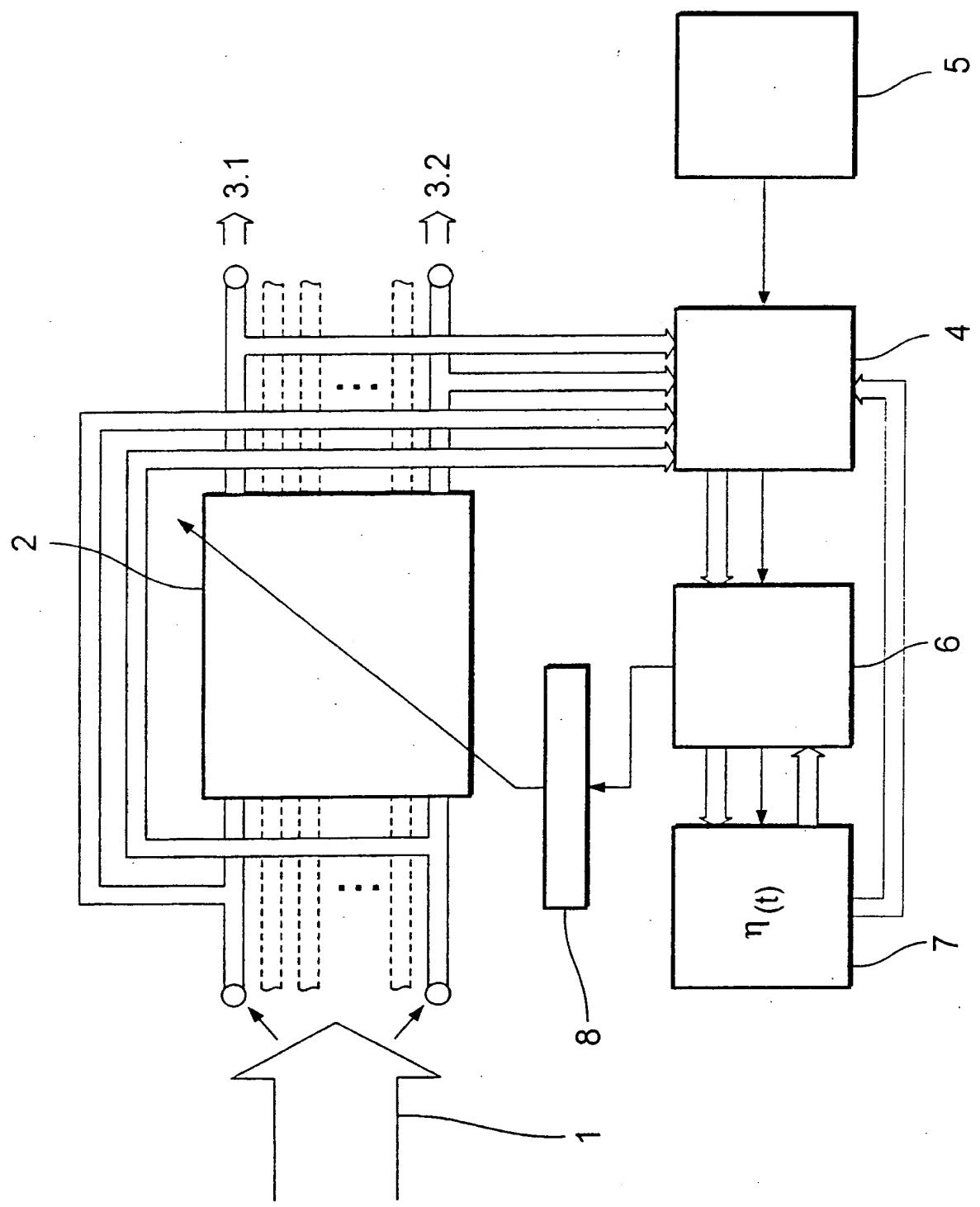


Fig.

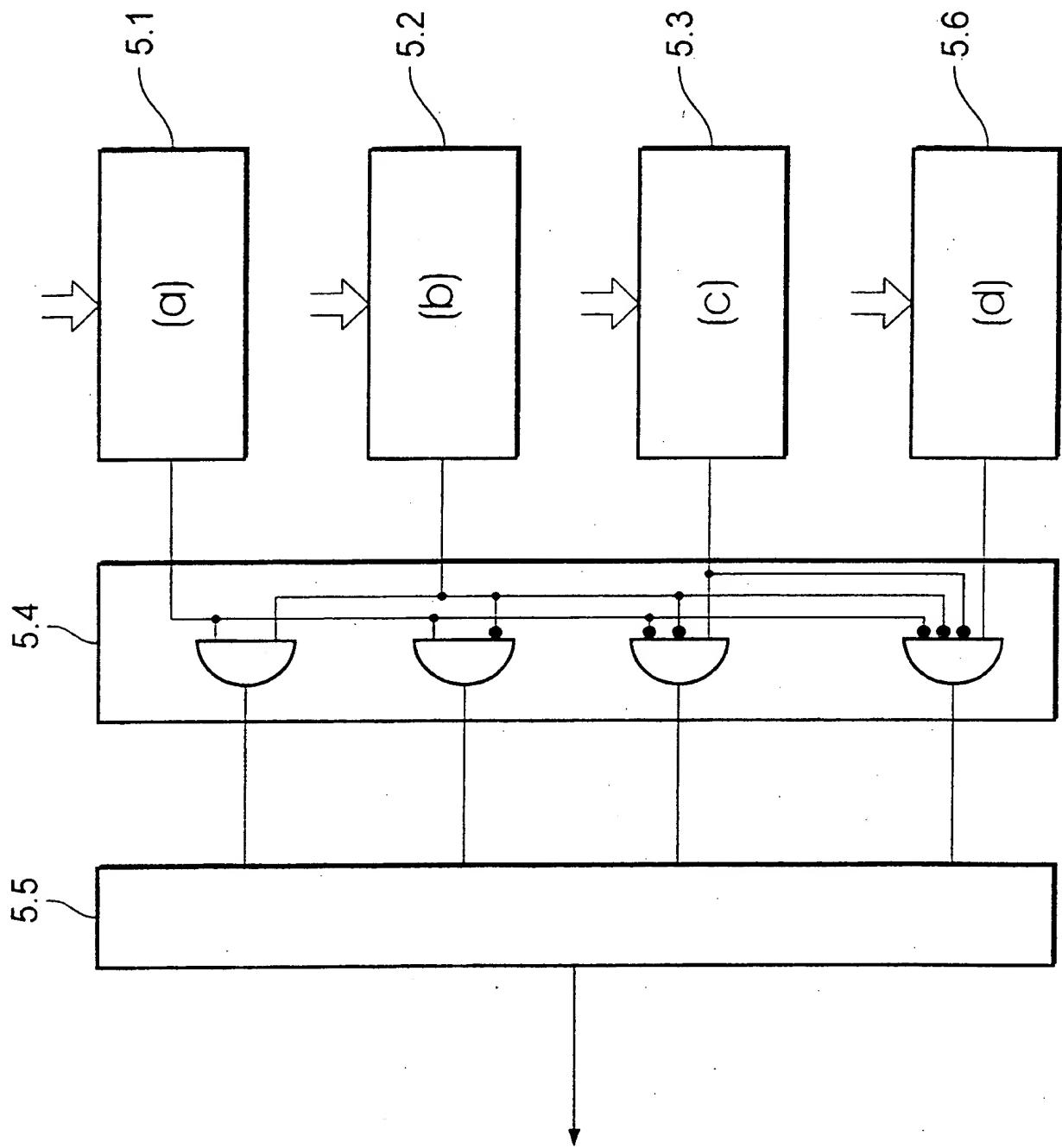


Fig. 1a

3/11

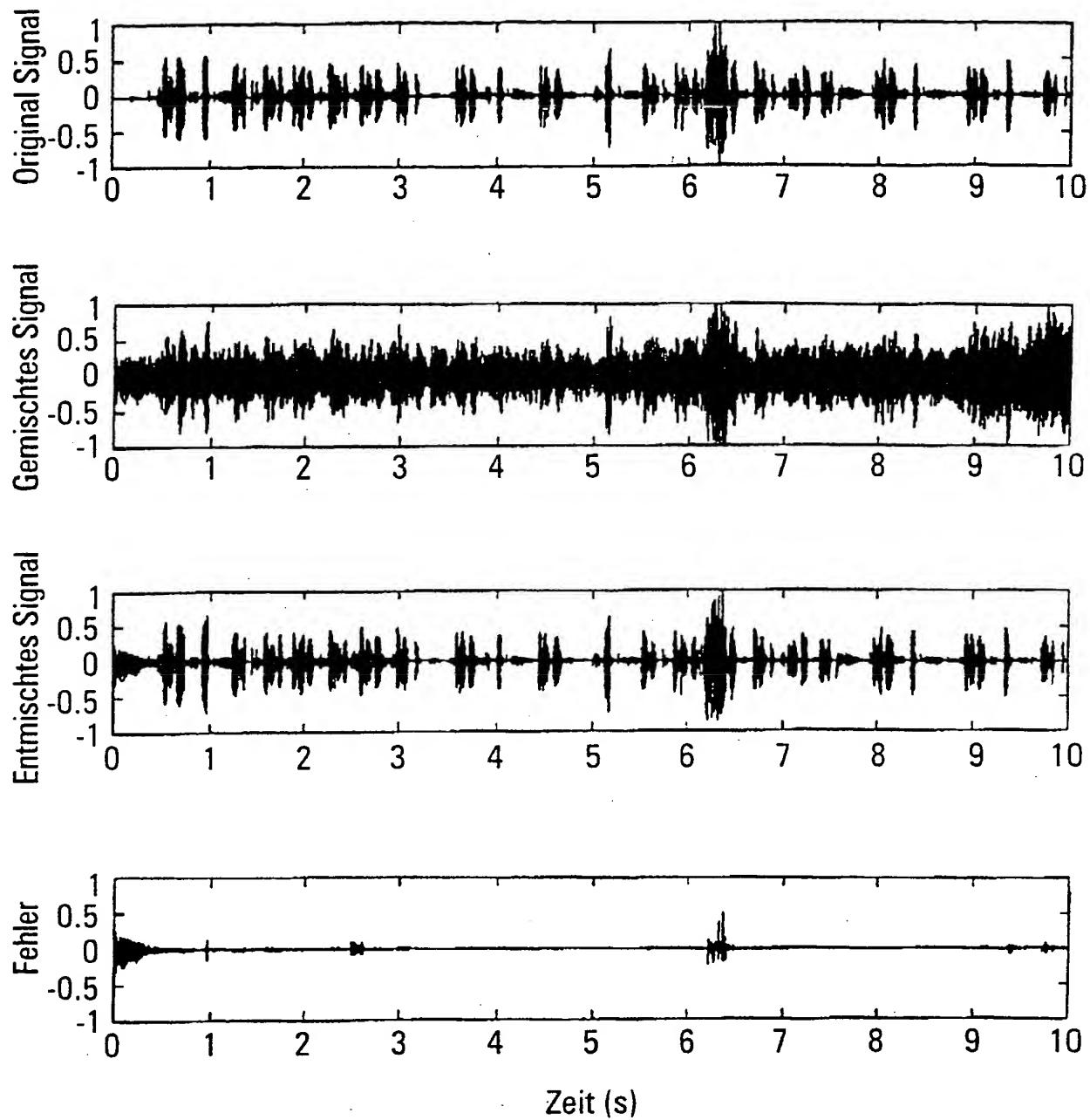


Fig.2

4/11

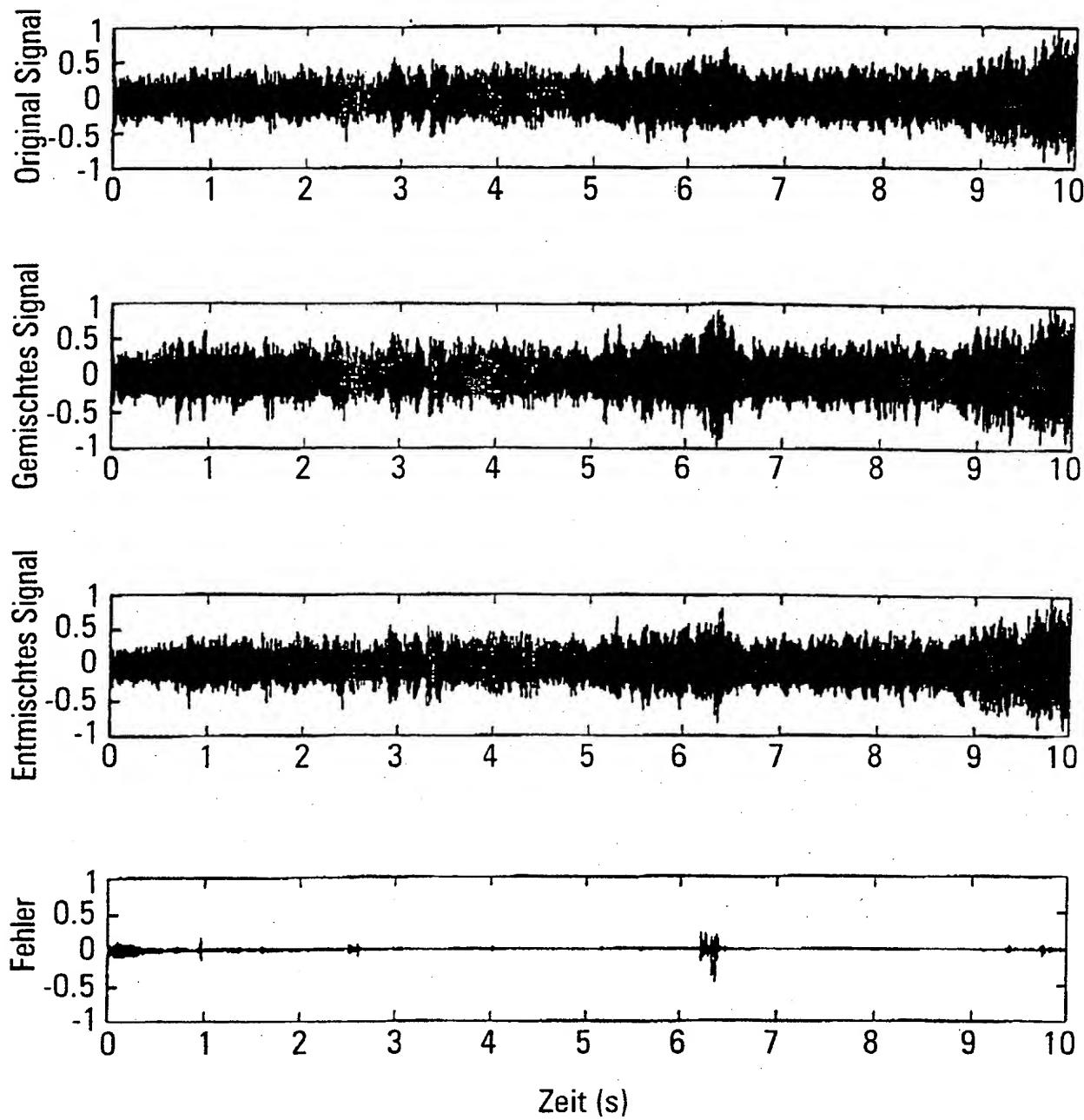


Fig.3

5/11

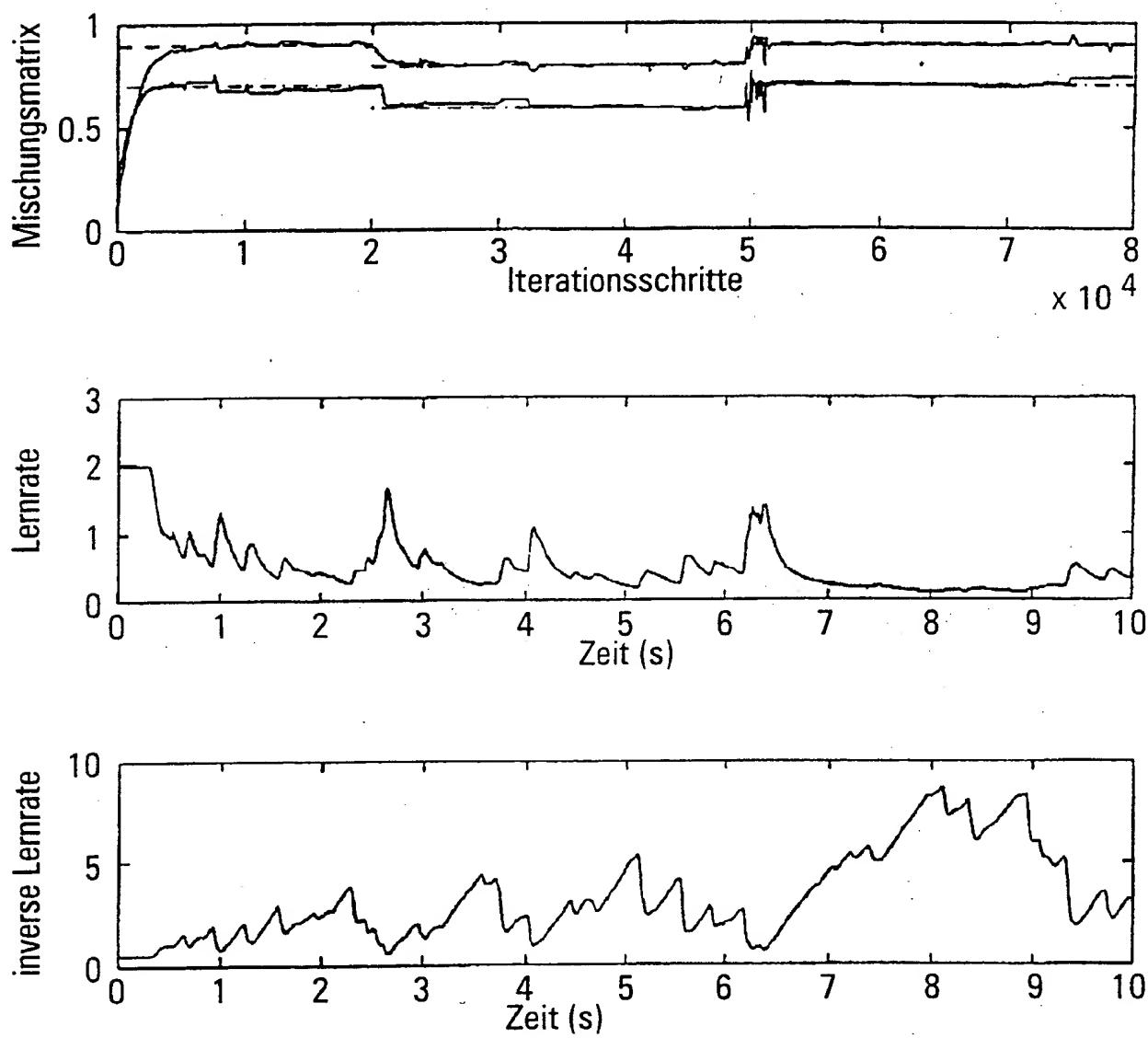


Fig.4

6/11

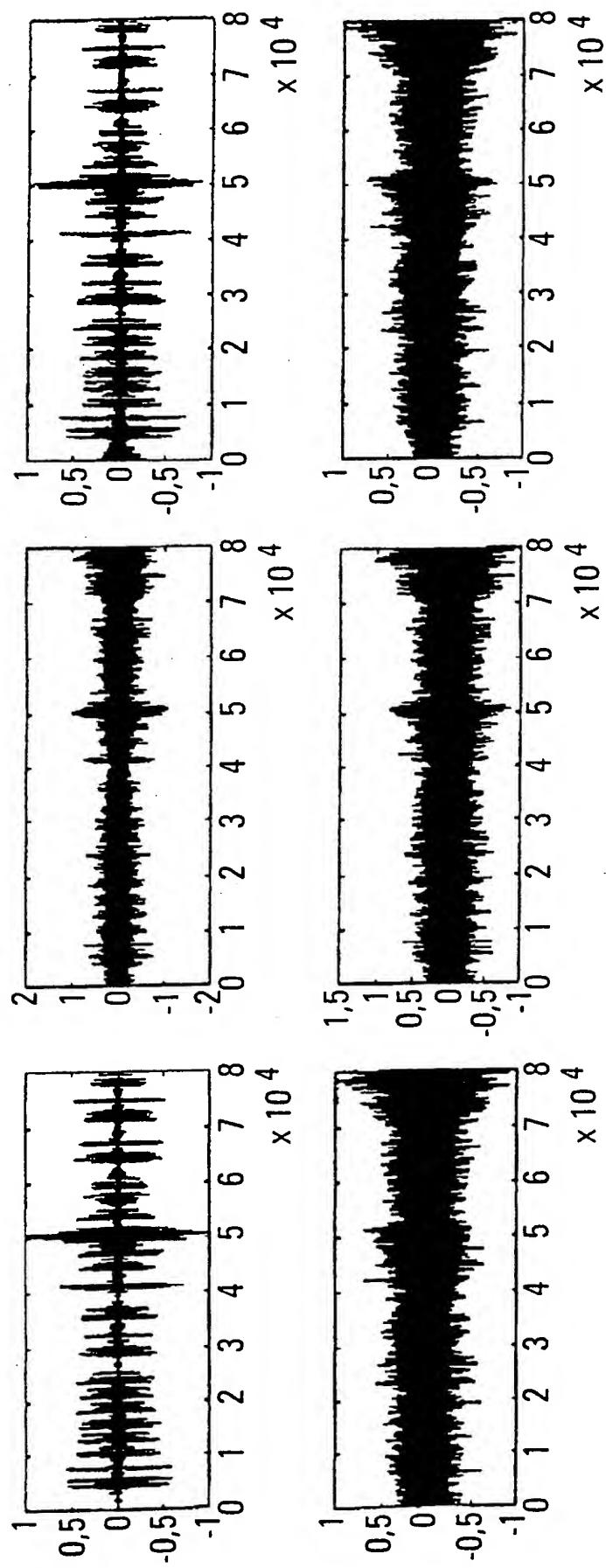


Fig.5

7/11

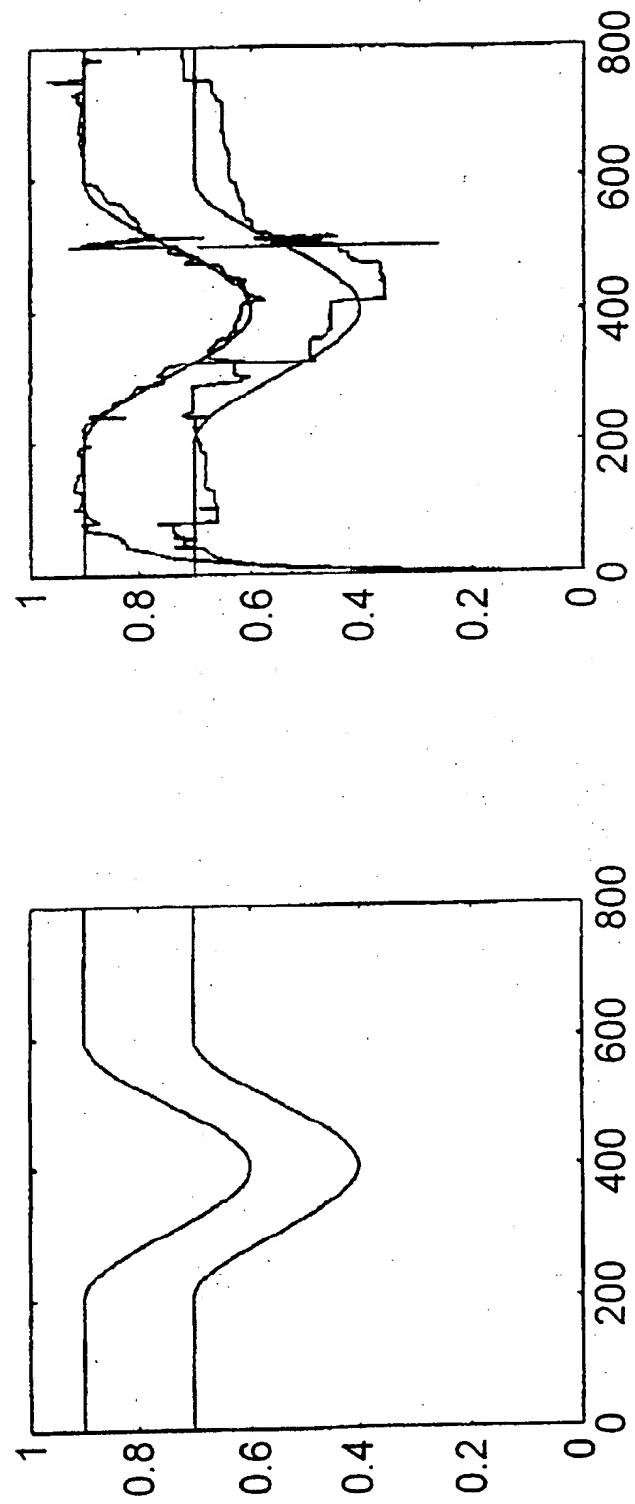


Fig.6

8/11

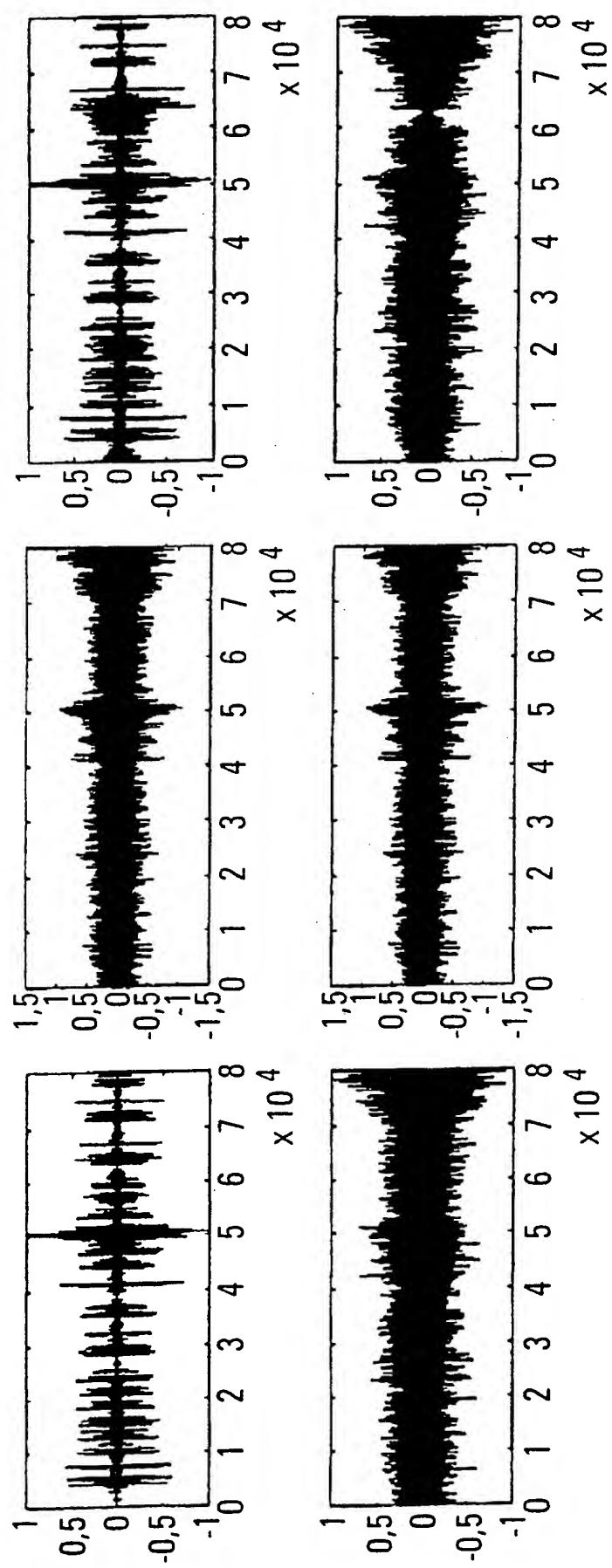


Fig.7

9/11

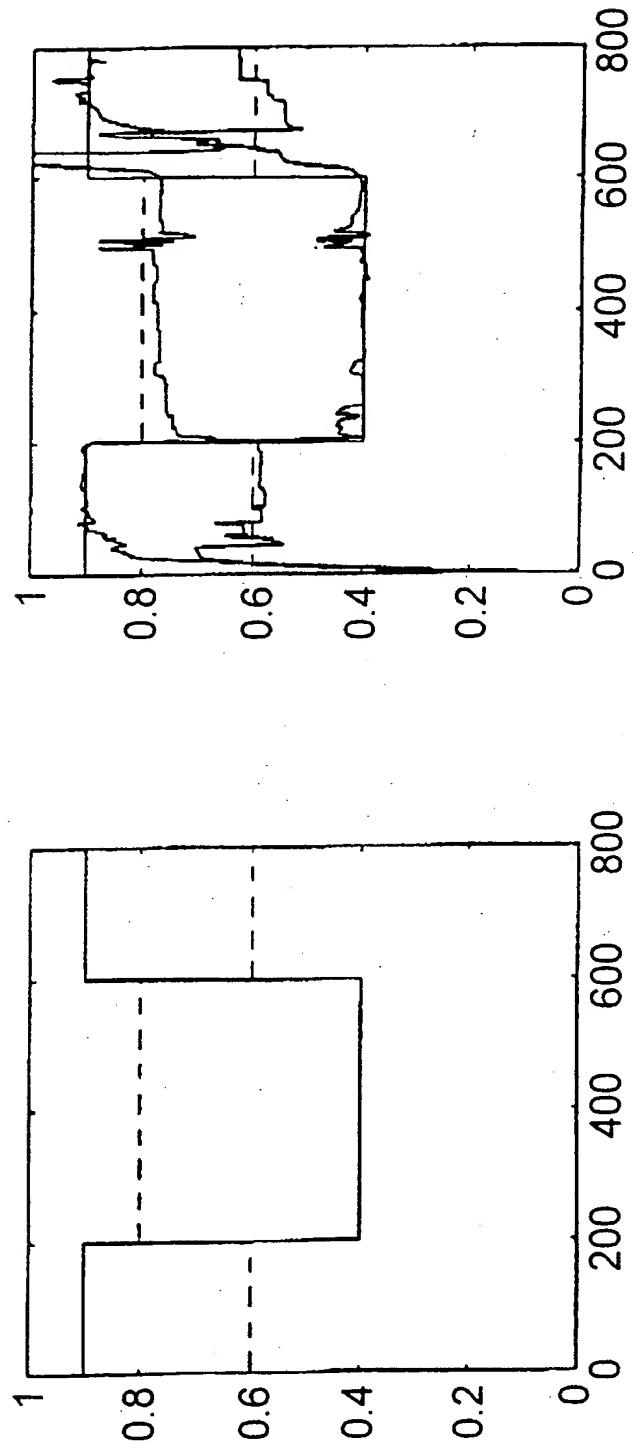


Fig.8

10/11

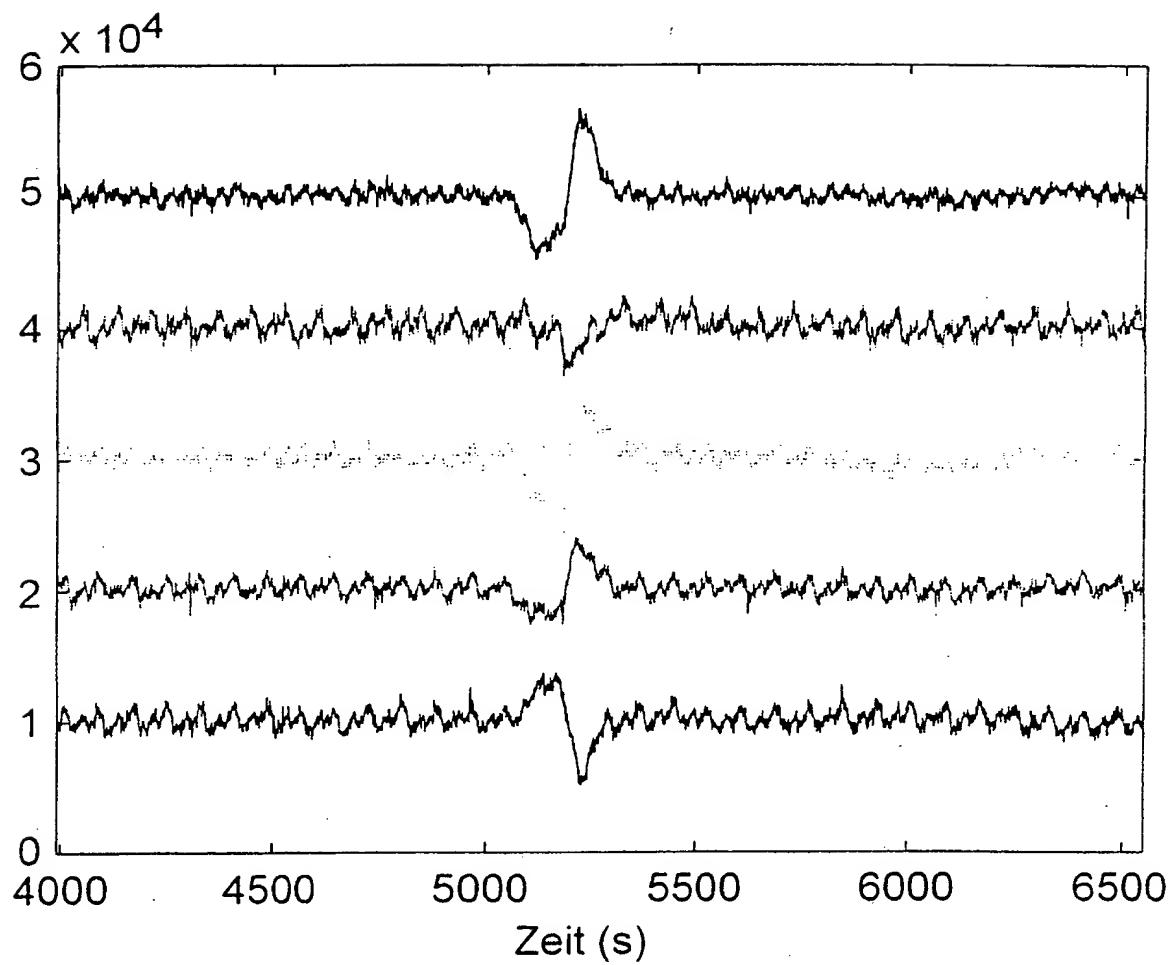


Fig.9

11/11

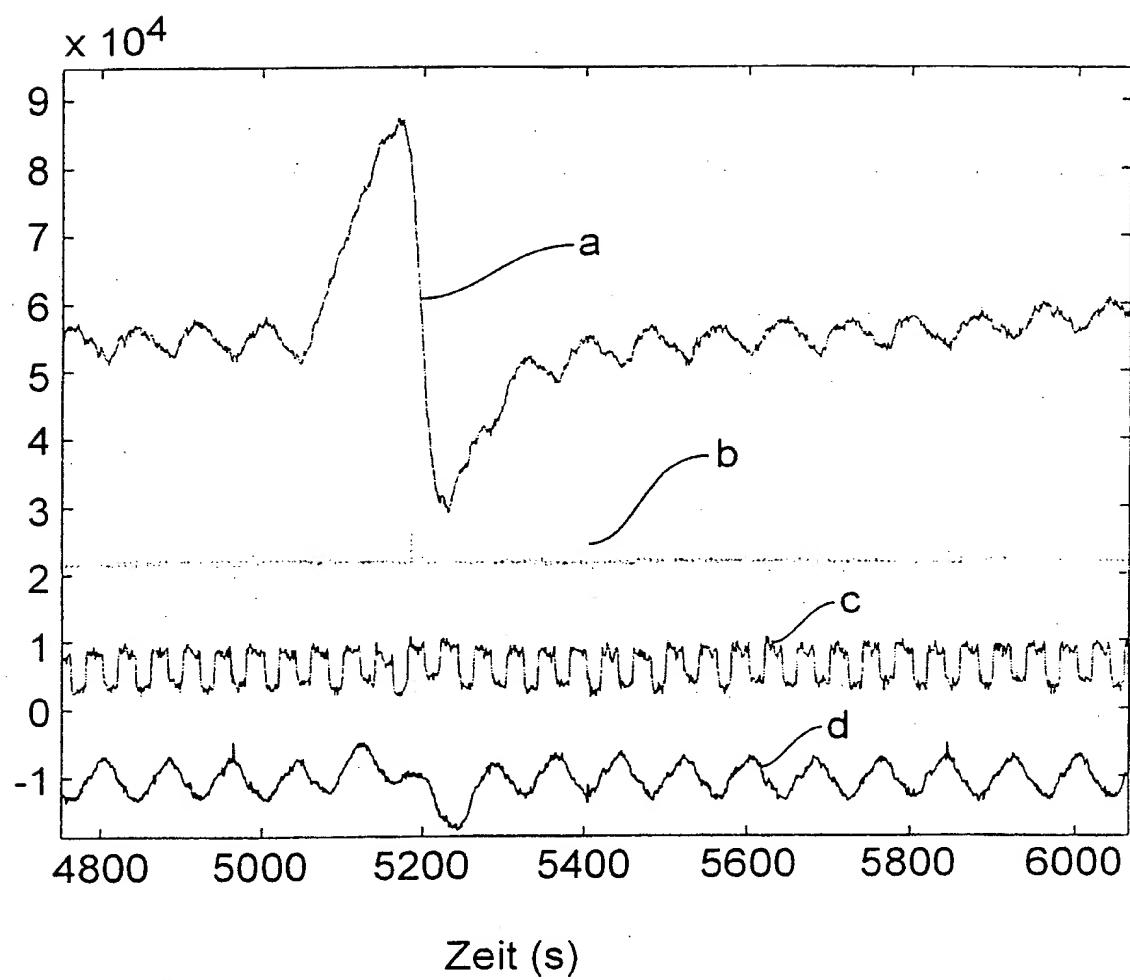


Fig.10

INTERNATIONAL SEARCH REPORT

International Application No
PCT/DE 97/02902

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F15/80

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MURATA N ET AL: "Adaptive online learning in changing environments" ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 9. PROCEEDINGS OF THE 1996 CONFERENCE, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 9. PROCEEDINGS OF THE 1996 CONFERENCE, DENVER, CO, USA, 2-5 DEC. 1996, ISBN 0-262-10065-7, 1997, LONDON, UK, MIT PRESS, UK, pages 599-605, XP002062107 see the whole document ---	1,2
A	-/-	3-10

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
14 April 1998	08/05/1998
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016	Authorized officer Schenkels, P

INTERNATIONAL SEARCH REPORT

International Application No

PCT/DE 97/02902

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VAN HULLE M M: "A novel unsupervised competitive learning rule with learning rate adaptation for noise cancelling and signal separation" NEURAL NETWORKS FOR SIGNAL PROCESSING IV. PROCEEDINGS OF THE 1994 IEEE WORKSHOP (CAT. NO.94TH0688-2), PROCEEDINGS OF IEEE WORKSHOP ON NEURAL NETWORKS FOR SIGNAL PROCESSING, ERMIONI, GREECE, 6-8 SEPT. 1994, ISBN 0-7803-2026-3, 1994, NEW YORK, NY, USA, IEEE, USA, pages 3-11, XP002062108 see page 3, line 1 - page 9, line 36; figures 1,2 ---	1
A	CICHOCKI A ET AL: "Self-adaptive neural networks for blind separation of sources" 1996 IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS. CIRCUITS AND SYSTEMS CONNECTING THE WORLD, ISCAS 96 (CAT. NO.96CH35876), 1996, ATLANTA, GA, USA, 12-15 MAY 1996, ISBN 0-7803-3073-0, 1996, NEW YORK, NY, USA, IEEE, USA, pages 157-160 vol.2, XP002062109 see page 157, left-hand column, line 1 - page 159, right-hand column, line 13 ---	1
A	CICHOCKI A ET AL: "ROBUST LEARNING ALGORITHM FOR BLIND SEPARATION OF SIGNALS" ELECTRONICS LETTERS, vol. 30, no. 17, 18 August 1994, page 1386/1387 XP000472070 see the whole document ---	1
A	BUREL G: "BLIND SEPARATION OF SOURCES: A NONLINEAR NEURAL ALGORITHM" NEURAL NETWORKS, vol. 5, no. 6, 1 November 1992, pages 937-947, XP000334414 see abstract see page 943, left-hand column, line 14 - right-hand column, line 21 -----	1

INTERNATIONALER RECHERCHENBERICHT

Intern. Aktenzeichen
PCT/DE 97/02902

A. KLASIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 6 G06F15/80

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
IPK 6 G06F

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie ³	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	MURATA N ET AL: "Adaptive online learning in changing environments" ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 9. PROCEEDINGS OF THE 1996 CONFERENCE, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 9. PROCEEDINGS OF THE 1996 CONFERENCE, DENVER, CO, USA, 2-5 DEC. 1996, ISBN 0-262-10065-7, 1997, LONDON, UK, MIT PRESS, UK, Seiten 599-605, XP002062107 siehe das ganze Dokument ---	1,2
A	-/-	3-10

Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

Siehe Anhang Patentfamilie

³ Besondere Kategorien von angegebenen Veröffentlichungen

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann nahelegend ist

"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

1

Datum des Abschlusses der internationalen Recherche

14. April 1998

Absendedatum des internationalen Recherchenberichts

08/05/1998

Name und Postanschrift der Internationalen Recherchenbehörde

Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Schenkels, P

INTERNATIONALER RECHERCHENBERICHT

Intern. Sales Aktenzeichen
PCT/DE 97/02902

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	VAN HULLE M M: "A novel unsupervised competitive learning rule with learning rate adaptation for noise cancelling and signal separation" NEURAL NETWORKS FOR SIGNAL PROCESSING IV. PROCEEDINGS OF THE 1994 IEEE WORKSHOP (CAT. NO.94TH0688-2), PROCEEDINGS OF IEEE WORKSHOP ON NEURAL NETWORKS FOR SIGNAL PROCESSING, ERMIONI, GREECE, 6-8 SEPT. 1994, ISBN 0-7803-2026-3, 1994, NEW YORK, NY, USA, IEEE, USA, Seiten 3-11, XP002062108 siehe Seite 3, Zeile 1 - Seite 9, Zeile 36; Abbildungen 1,2 ---	1
A	CICHOCKI A ET AL: "Self-adaptive neural networks for blind separation of sources" 1996 IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS. CIRCUITS AND SYSTEMS CONNECTING THE WORLD, ISCAS 96 (CAT. NO.96CH35876), 1996, ATLANTA, GA, USA, 12-15 MAY 1996, ISBN 0-7803-3073-0, 1996, NEW YORK, NY, USA, IEEE, USA, Seiten 157-160 vol.2, XP002062109 siehe Seite 157, linke Spalte, Zeile 1 - Seite 159, rechte Spalte, Zeile 13 ---	1
A	CICHOCKI A ET AL: "ROBUST LEARNING ALGORITHM FOR BLIND SEPARATION OF SIGNALS" ELECTRONICS LETTERS, Bd. 30, Nr. 17, 18.August 1994, Seite 1386/1387 XP000472070 siehe das ganze Dokument ---	1
A	BUREL G: "BLIND SEPARATION OF SOURCES: A NONLINEAR NEURAL ALGORITHM" NEURAL NETWORKS, Bd. 5, Nr. 6, 1.November 1992, Seiten 937-947, XP000334414 siehe Zusammenfassung siehe Seite 943, linke Spalte, Zeile 14 - rechte Spalte, Zeile 21 -----	1